



AFRL-RH-WP-TR-2013-0067

WARFIGHTER VISUALIZATIONS COMPILATIONS

**Kristen K. Liggett, Ph.D.
Battlespace Visualization Branch**

**MAY 2013
Interim Report**

Approved for public release; distribution is unlimited.

See additional restrictions described on inside pages

**AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC).

AFRL-RH-WP-TR-2013-0067 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//
KRISTEN K. LIGGETT, PH.D.
Warfighter Visualizations Workunit Manager
Battlespace Visualization Branch

//signed//
JEFFREY L. CRAIG
Chief, Battlespace Visualization Branch
Warfighter Interface Division

//signed//
WILLIAM E. RUSSELL
Acting Chief
Warfighter Interface Division
Human Effectiveness Directorate
711 Human Performance Wing

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.</small>				
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 15/05/2013		2. REPORT TYPE Interim		3. DATES COVERED (From - To) Jan 2011 - Dec 2012
4. TITLE AND SUBTITLE WARFIGHTER VISUALIZATION COMPILATIONS			5a. CONTRACT NUMBER IN-HOUSE	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER 62202F	
6. AUTHOR(S) Kristen K. Liggett, PhD			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER (H03E) 71841142	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Battlespace Visualization Branch 2255 H Street Wright-Patterson AFB, OH 45433			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711 Human Performance Wing Human Effectiveness Directorate Warfighter Interface Division Battlespace Visualization Branch Wright-Patterson AFB OH 45433			10. SPONSOR/MONITOR'S ACRONYM(S) 711 HPW/RHCV	
			11. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-RH-WP-TR-2013-0067	
12. DISTRIBUTION AVAILABILITY STATEMENT Distribution A: Approved for public release; distribution unlimited.				
13. SUPPLEMENTARY NOTES Various Clearance Numbers (see footer and list of publications in report)				
14. ABSTRACT This interim report contains a compilation of nine publications generated from research conducted under the Warfighter Visualization workunit over the past 2 years. The workunit objectives are to evaluate 3D display technologies, evaluate wireless interaction devices for effective use with 3D displays, and explore visual analytics techniques and extend them to Air Force applications. As can be seen by the collection of publications, work in 3D displays has continued during this period and the visual analytics techniques investigated over the past two years were applied to social and cyber networks.				
15. SUBJECT TERMS Three-dimensional (3D) displays, visual analytics, human factors evaluations, social network visualization, cyber visualization, Visual Analytics Science and Technology (VAST)				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	SAR	73
			19a. NAME OF RESPONSIBLE PERSON Kristen K. Liggett, Ph.D.	
			19b. TELEPHONE NUMBER (Include area code)	

THIS PAGE INTENTIONALLY LEFT BLANK.

Acknowledgements

I would like acknowledge the following authors who contributed to the papers contained in this interim report:

Erica Edelman – Wright State University

John McIntire – 711 HPW/RHCV

O. Isaac Osesina – University of Arkansas at Little Rock

Michael Craft – Middletown High School

Paul Havig – 711 HPW/RHCV

Eric Geiselman – 711 HPW/RHCV

Eric Heft – 711 HPW/RHCV

Cecilia Bartley – University of Arkansas at Little Rock

M. Eduard Tudoreanu – University of Arkansas at Little Rock

Georges Grinstein – University of Massachusetts Lowell

Kristin Cook – Pacific Northwest National Laboratory

Bohdan Nebesh – Department of Defense

Mark Whiting – Pacific Northwest National Laboratory

Kirsten Whitley – Department of Defense

Shawn Konenci – University of Massachusetts Lowell

Fairul Mohd-Zaid – 711 HPW/RHCV

Michael Cooper – Pacific Northwest National Laboratory

Celeste Lyn Paul – National Security Agency

List of Publications

1. Layered Interactive Visual Interface Design: A Visual Interface for the Navigation and Analysis of Digital Text Communications by Erica Edelman and John McIntire. Cleared for public release on 07 Jan 2011, 88ABW-2011-0048.
2. Development of Visualizations for Social Network Analysis of Chatroom Text by John McIntire, O. Isaac Osesina, and Michael Craft. Cleared for public release on 07 Jan 2011, 88ABW-2011-0047.
3. How Much Camera Separation Should Be Used for the Capture and Presentation of 3D Stereoscopic Imagery on Binocular HMDs? By John McIntire, Paul Havig, Eric Geiselman, and Eric Heft. Cleared for public release on 10 Mar 2011, 88ABW-2011-1269.
4. Visualizing Weighted Networks: A Performance Comparison of Adjacency Matrices versus Node-Link Diagrams by John P. McIntire, O. Isaac Osesina, Cecilia Bartley, M. Eduard Tudoreanu, Paul R. Havig, and Eric E. Geiselman. Cleared for public release on 12 Oct 2011, 88ABW-2011-5408.
5. What is 3D Good For? A Review of Human Performance on Stereoscopic 3D Displays by John P. McIntire, Paul R. Havig, and Eric E. Geiselman. Cleared for public release on 12 Oct 2011, 88ABW-2011-5412.
6. VAST 2011 Challenge: Cyber Security and Epidemic by Georges Grinstein, Kristin Cook, Paul Havig, Kristen Liggett, Bohdan Nebesh, Mark Whiting, Kirsten Whitley, and Shown Konenci. Cleared for public release on 05 Dec 2011, 88ABW-2011-6206.
7. Methods for Extracting Social Network Data for Chatroom Logs by O. Isaac Osesina, John P. McIntire, Paul R. Havig, Eric E. Geiselman, Cecilia Bartley, and M. Eduard Tudoreanu. Cleared for public release on 27 Mar 2012, 88ABW-2012-1741.
8. Why Social Network Analysis is Important to Air Force Applications by Paul R. Havig, John P. McIntire, Eric Geiselman, and Fairul Mohd-Zaid. Cleared for public release on 29 Mar 2012, 88ABW-2012-1853.
9. VAST Challenge 2012: Visual Analytics for Big Data by Kristin Cook, Georges Grinstein, Mark Whiting, Michael Cooper, Paul Havig, Kristen Liggett, Bohdan Nebesh, and Celeste Lyn Paul. Cleared for public release on 10 Oct 2012, 88AWB-2012-5370.

Layered Interactive Visual Interface Design: A Visual Interface for the Navigation and Analysis of Digital Text Communications

Erica Edelman
Wright State University
Dayton, OH
edelman.4@wright.edu

John McIntire
711th Human Performance Wing / RHCvZ
Air Force Research Laboratory, WPAFB, OH
john.mcintire@wpafb.af.mil

POSTER PAPER

ABSTRACT

We describe the preliminary development of an interactive visualization tool intended to produce a multi-layered visual interface design that allows for fast, easy navigation and analysis of digital text corpora and text communications, allowing for more efficient distributed collaboration and communications in the virtual world.

KEYWORDS: information visualization, text communications, sentiment analysis, visual interface design, human-computer interaction, internet navigation.

1. INTRODUCTION

One of the weaknesses of current information technology is that it generates and presents tremendous amounts of data, but rarely manages to present both high-level overviews and low-level details of that data in a method that is intuitively understandable to a human consumer. The practice of information visualization has been applied very effectively to multiple fields of need including: sensor, scientific, medical, and financial data, and internet communications. When dealing with the latter, however, the sheer enormity of the available information is difficult for even the best tools or visualizations to convey the many intricate patterns and meanings therein. This is particularly true for text-based communications and for digital text corpora. How, then, can the vast deposit of data that can be accessed via the internet (or other digital text data) be presented in a manner that is both intuitive and useful to the information consumer and digital communicator?

Although a single visualization, while useful, may not be up to such a prodigious task, an interface consisting of multiple layers of a variety of visualizations could perhaps provide enhanced usability, rapid understanding, and more intuitive interaction. For this reason, a joint team of researchers developed the Layered Interactive Visual Interface Design, or LIVID. This work is meant to show how such an interface could be developed, to explore potentially fruitful ideas in this area, and to provide a starting point for future researchers, rather than serve as an end product itself.

LIVID uses, at its core, basic text analysis techniques to present text-based internet content or digital text corpora in a more easily-navigable and understandable manner for the everyday user, while also providing useful interactive visualizations of potential interest to researchers. Many of its envisioned visualization methods are not new (comprehensive reviews of visualizations and systems for analyzing/using online digital text can be found in [1,2]). But some (such as the WordNets) we believe are unique and may be particularly useful extensions to word-cloud-type visualizations. Additionally, this tool is specifically intended to help regular users engage with and collaborate more efficiently and effectively with their online communities and virtual worlds.

2. LIVID

2.1. Overview

LIVID consists of a series of “layers”, each representing a different zoom magnification of the internet (or groupings of information sources) as a whole. The outermost layer

displays all of the user's particular text sources or "channels" of interest at once, while providing information that allows the user to effectively decide whether or not to examine each channel individually. This is much like a "dashboard" concept. For the purposes of our prototype, this outer layer contains the user's favorite websites and digital text sources. The middle layer provides visualization methods for each singular source. The outer and middle layers fulfill the first two rules of Ben Shneiderman's Information-Seeking Mantra: first, provide an overview of all the data; then, provide the ability to do some preliminary zooming and filtering by the user [1]. The innermost layer displays a single webpage, however it performs slight alterations to the page to make it easier to find the relevant information; this fulfills Shneiderman's third rule: provide the user with easy access to detailed information on demand.

2.2. Outer Layer

The outer layer (Figure 1) displays a list of the user's favorite websites or other textual content, sub-categorized into types, such as blogs, social networking sites, comics, videos, or news sources. We envision our user to be a regular communicator and collaborator within the virtual world. Much internet traffic is due to checking a website solely to see if there is anything new posted on it. LIVID gets rid of this need: if a page has been updated since the last visit, that page is highlighted and moved to the top of the category. Instead of checking 15 pages, or getting 15 emails about new updates, a single glance at LIVID's outer layer should provide a quick reference about communications or page status.

However, this is not the only method the outer layer utilizes in assisting the user in deciding whether or not to visit a page or text document: A word cloud is also visible for each page that shows relevant information of the user's preference. For Hulu.com, a website that streams previously aired TV shows and movies, the user might want to know what shows have been added to their queue. For a blog, the user might want to know what topics are being discussed or the titles of the most recent blog postings. For an online newspaper, the user may be interested in the titles of the five most popular articles of the day. If desired, the words in the clouds can be color coded to relay additional information.

A straightforward example would be to do a positive/negative affective scale, where blog posts and videos that receive highly positive comments are yellow, while pages that receive generally negative comments are blue. Those that fall in the middle are green. Certain pages, such as webcomics, would not result in useful word clouds. For these, other options are available: The example in the

prototype shows a random archived comic from the website. Other options include thumbnail strips of imagery or dynamic snippets of videos.

Where the word cloud contains explicit page names (i.e. article titles or video names) clicking on that name will navigate the user directly to the desired page. Otherwise, the outer layer also contains a thumbnail of the webpage's main page that expands when rolled over by the mouse cursor. Clicking on the thumbnail will direct the user to the webpage's main page. The middle layer is accessed by clicking on the website name.

2.3. Middle Layers

The middle layers consist of visualizations for the selected website. Due to the individualized nature of web pages, different visualization methods work best on different sites. Therefore these middle layers have to be highly customizable. The current LIVID design does not include an interface or method for customization, which is a factor future work can examine.

The example shown in Figure 2 is a visualization for Twitter which was developed specifically for this project, based loosely on the idea of Havre and colleagues' ThemeRiverTM [2]. The map on the upper-right shows which areas have positive or negative posts. The yellow areas have an abundance of positive/happy posts, while the blue areas lean towards negative/unhappy posts. This usage of yellow-blue coding on the positive-negative scale is used throughout LIVID's navigation. To the left of the map is a stack graph showing posting amount by state over time. It is also yellow-blue color coded.

Beneath the state stack graph is another stack graph that is instead broken down by topic. To the right of the topic stack graph is an enlargeable color-coded word cloud of twitter topics. Fully zoomed out, it shows the main twitter topics such as Entertainment, Politics, Humor, etc. The size of the word is coded to amount of tweets. As the user zooms in on an area, sub-categories fade in and appear. Further zooming causes sub-categories of the sub-categories to appear, and so forth. For example, zooming in on "Entertainment" would show sub-categories such as "Media", "Sports", "Hobbies", etc. Zooming in on "Media" would reveal "Games", "Television", "Movies", "Books", and so forth.

These visualizations can be animated by pressing the play button. The map and word cloud animate via the variable time, while the red line on the stack graphs scrolls across to show where that specific time is located on the x-axis. The animation can be paused, and the red lines can be dragged on the graph to show a specific day. An example

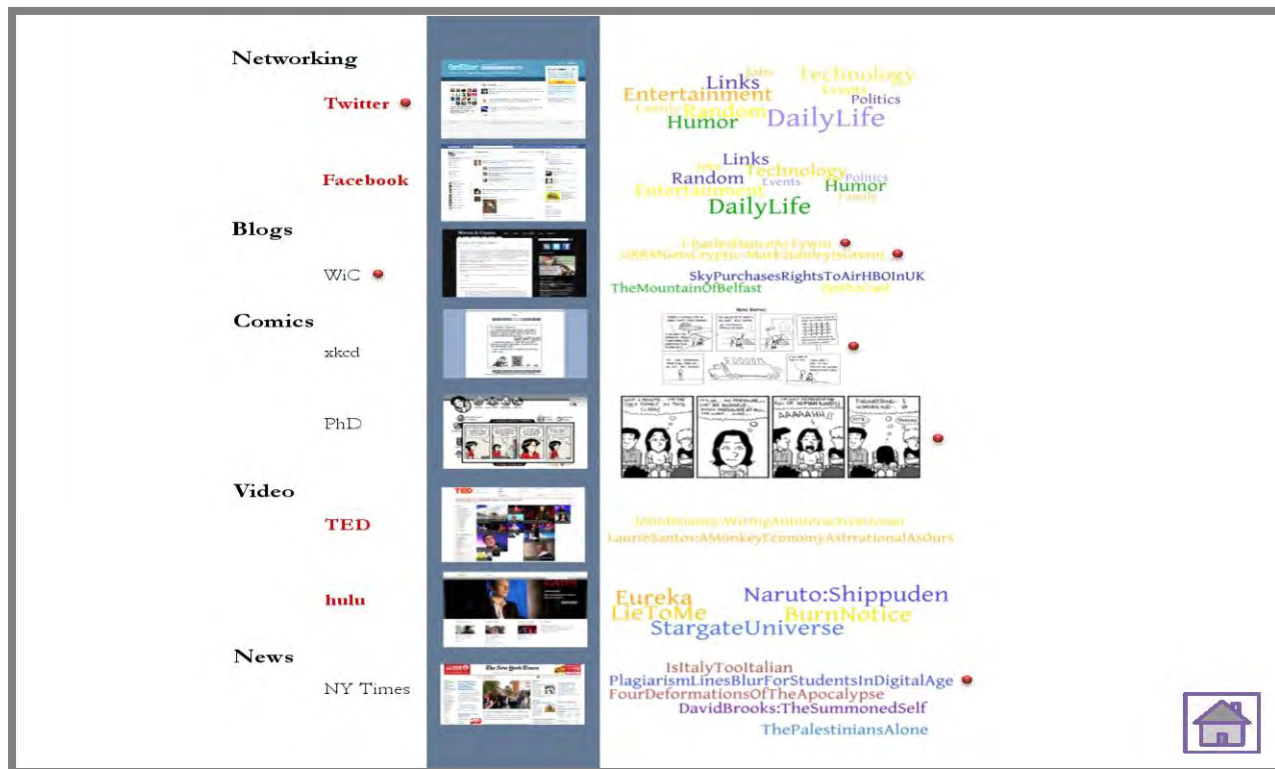


Figure 1. LIVID's "Outer Layer" which Allows Users to Quickly View Sites/Content and Updates

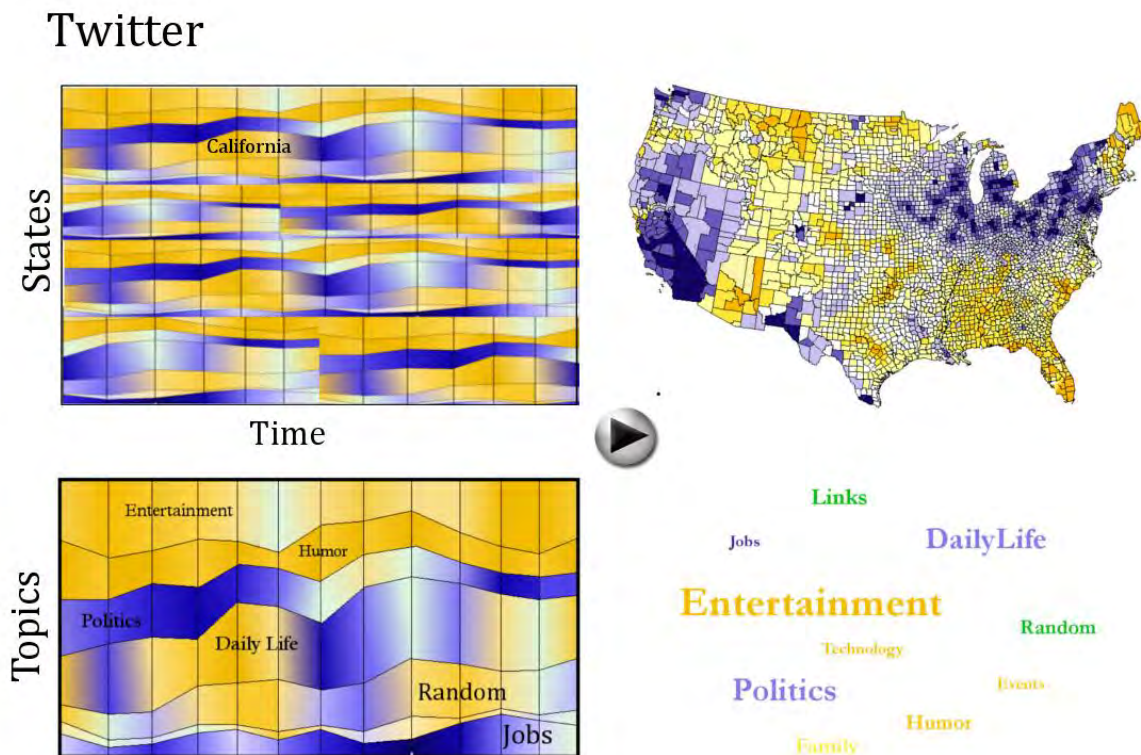


Figure 2. Twitter Visualizations across Topics, Locations, and Time (a Middle Layer)

of how this page would be altered if using a different web site for a base is shown in Figure 3. This page draws its data from a fan blog. It utilizes three visualizations. The first is a word cloud that displays the most popular words since the user's last visit.

The second visualization is a graph developed specifically for LIVID: The green line represents number of comments over time, while the purple line represents blog posts per time. The x-axis is time, while the y-axis is number of posts, normalized to 0-1. Areas where the green line is above the purple line are colored green. This shows where there was a relatively high amount of discussion per post. Areas where the purple line is above the green line are colored purple, and show where there was not a lot of discussion per post. When the graph is zoomed in, words appear in the colored areas that represent the topics that are most likely responsible for the popularity of the posts. These are found by looking at the words that appear frequently in the green area that tend not to appear in purple areas, and vice versa. For example, Figure 3 shows that people comment a lot on blog posts relating to HBO, Casting, and Clues, but tend not to comment on blog posts relating to Ireland and Sets.

The final visualization on this page is a word tree. This is an interactive chart in which the user can enter a word and see all the times the word appeared on the website, along with the proceeding or following content. This is an exceptional way to search a site for particular content, because it shows the word in context. If the user sees a sentence they would like to navigate to, they can click on the sentence and be directed to that location.

2.4. Inner Layers

The inner layers consist of individual web-pages, or what the user would normally expect to see on their internet browser. However, there are many ways an interface can make consuming this data quicker and easier, two of which were developed for LIVID and presented below:

Many websites such as news sites, blogs, social networking sites, or video sharing sites offer their users the ability to comment on their content. For popular sites or posters, the number of comments can reach well into the hundreds and even thousands. The above figure (Figure 4) shows a method of marking these comments to allow a reader to get the relevant information from them without having to spend a vast amount of time reading through them all.

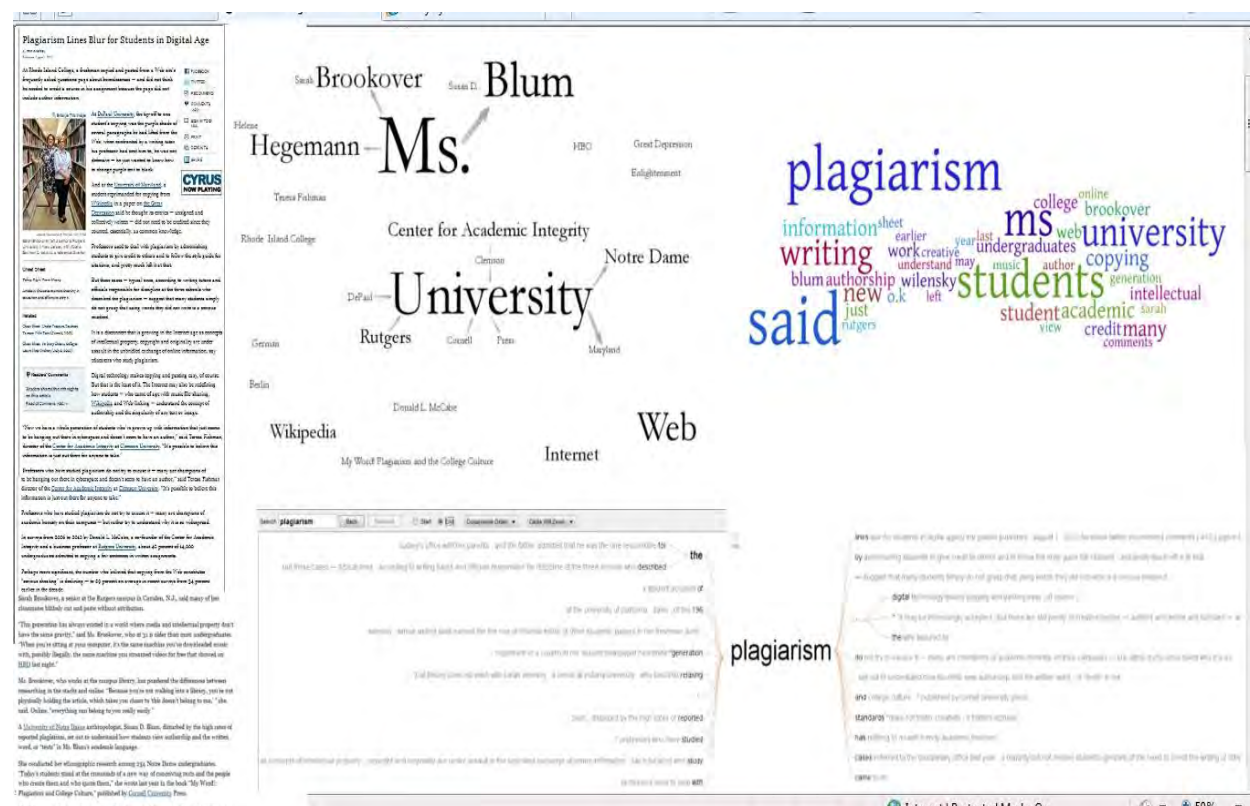
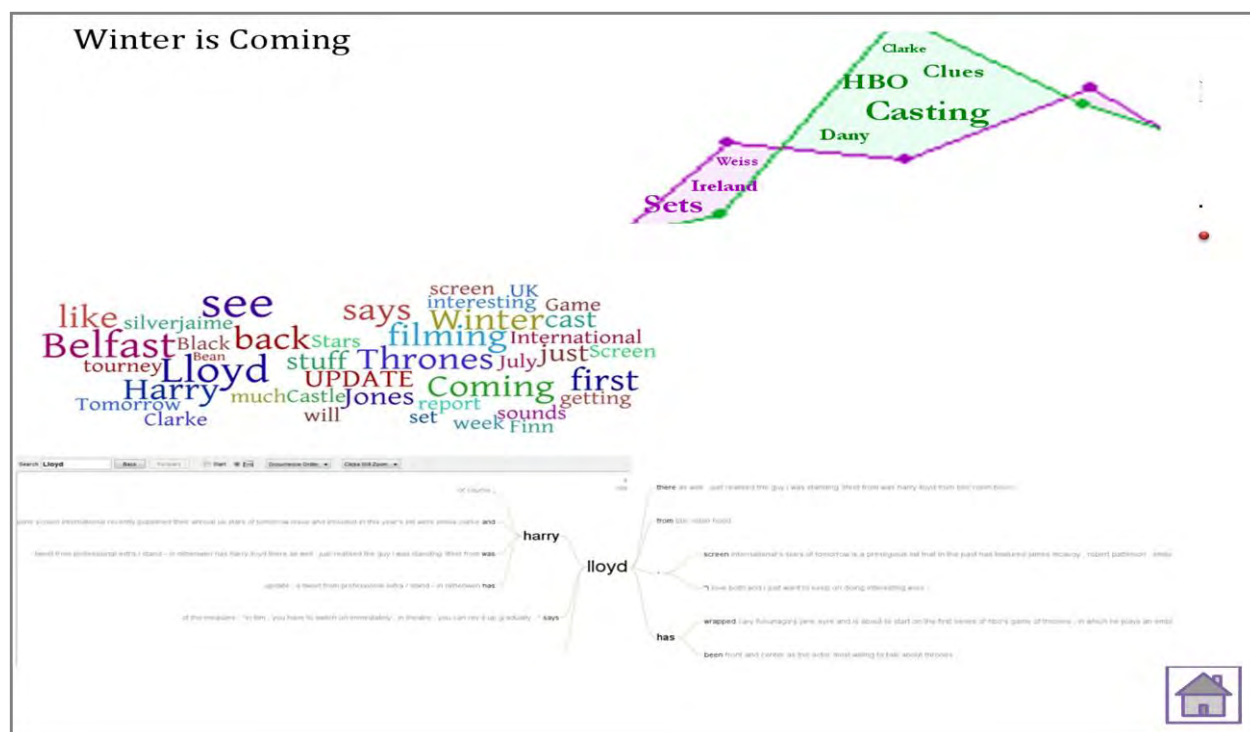
The first thing an internet user might come to realize about comments is that many are "throwaway" comments whose sole purpose is to make the author's presence

known. Other comments exist primarily to convey a positive or negative sentiment about the issue at hand. Using html tags, the background of positive comments is shaded yellow, while the background of negative comments is shaded blue. Next, using relevancy analysis, a text analyzer should be able to tell which comments are very relevant, and which are "throwaway". The most irrelevant comments can be made increasingly transparent so that they blend into the background, while the most relevant and important comments are made bigger, bold, and surrounded by a red border. Furthermore, when topics are changed, comments can be made indented or otherwise visually indicated. Therefore, instead of reading through hundreds of comments, a user can scroll down looking for the big red boxes and reading the contained text. As they scroll, they can note the colors to see if general positive or negative sentiment is expressed. This gives the user the general data that is normally the reason for reading comments in the first place.

The next example shows how a LIVID-type system can improve a news article page. The user can choose whether to maximize either the article page, or its visual representation. This consists of three visualizations: a Proper Noun Word Net, a word cloud, and a navigable word tree.

The Proper Noun Word Net was developed for LIVID, and is useful for larger pieces of text (Figure 4). Proper Nouns are found by looking for capitalized words that are not the first word of the sentence, then going back and seeing if any first-words match the found proper nouns, and including them as well if they do. However many proper nouns are compound and consist of multiple words, for example Lt. Bob Smith, The Great Depression, or Northern Ireland. Therefore, the analyzer then looks to find pronouns that occur next to each other. If the two occur exclusively with each other (for example if every time the word "Great" appears in the article, the word "Depression" follows, and vice versa) then they are considered to be one unit. If they occur together but not exclusively (for example if the article mentions "Wright State University" and "Ohio State University"), then they are linked via an arrow, forming a "net". Number of mentions is coded to size of the word.

This visualization is useful in multiple ways. Firstly, it answers the specifics of "who, what, and where" for a given piece of text. Secondly, via the net function, the user can easily find relationships among specific entities. For example, they can see all the titles (Dr., Mrs., General, etc) in a piece of text and see who holds those titles. Via matching last names (Bob Smith and Jane Smith) familial relationships can be found. Links can also be found in categories and sub-categories, such as: Dayton, Ohio and



Cincinnati, Ohio; The Eiffel Tower and the Leaning Tower of Pisa; or Wright-Patterson Air Force Base and Maxwell Air Force Base. In Figure 5, the user can see that many different Universities are discussed.

3. CONCLUSIONS AND FUTURE WORK

LIVID is a research prototype visualization system meant to explore potentially fruitful areas in information visualization, navigation, and interaction, and to spark future research and ideas. At this point, the interface is still predominantly in the idea-generation-and-testing phase. Future work could include creating a working model of a LIVID-style interface, or of a completely new interface design. This interface would need to add further interactive functionality, such as the ability to personalize the layout and structure of the interface; or to tailor it to more specific applications if desired. We would like future iterations of LIVID to possess the ability to include humans-in-the-loop so that the interface itself is dynamic, responsive, and adaptable to users and their unique needs and behaviors.

ACKNOWLEDGEMENTS

We wish to thank the Tec^Edge Summer at the Edge (SATE) program for its invaluable support, particularly that provided by Dr. Rob Williams from SATE. Special thanks go to Isaac Osesina (University of Arkansas Little Rock) and Michael Craft (Middletown High School), both fellow SATE members. Thanks to the Air Force Research Laboratory's 711th Human Performance Wing Information Visualization section for their support, particularly from mentor Eric Geiselman. Thanks also to Dr. Steven Rogers at the Air Force Research Laboratory and Adam Rogers at Wright State University for their many helpful and insightful discussions and ideas. This work was primarily conducted by the Information Visualization Team at Tec^Edge's Summer at the Edge student program, under the guidance of the Air Force Research Laboratory's 711th Human Performance Wing Information Visualization section, over 10 weeks in the summer of 2010.

REFERENCES

- [1] Viegas, F., "Revealing individual and collective pasts: Visualizations of online social archives," Ph.D. dissertation in Media Arts and Sciences at the Massachusetts Institute of Technology, 2005. Available online at: http://smg.media.mit.edu/papers/Viegas/viegas_thesis.pdf.
- [2] Silic, A., "Visualizatin of text streams: A survey," Proceedings of the 14th International Conference on Knowledge-based and Intelligent Information and Engineering Systems: Part II, 2010.
- [3] Shneiderman, B. "The eyes have it: A task by data type taxonomy for information visualizations," Proceedings of the IEEE Symposium on Visual Languages, pp. 336-343, 1996.
- [4] Havre, S., Hetzler, E., Whitney, P., & Nowell, L. "ThemeRiver: Visualizing thematic changes in large document collections," IEEE Transactions on Visualizations and Computer Graphics, vol. 1, issue 8, pp. 9-20, 2002.

Development of Visualizations for Social Network Analysis of Chatroom Text

John McIntire
Air Force Research Laboratory
Wright-Patterson AFB, OH
john.mcintire@wpafb.af.mil

O. Isaac Osesina
University of Arkansas at Little Rock
Little Rock, AR
oioosesina@ualr.edu

Michael Craft
Middletown High School
Middletown, OH
mcraft@peak15.org

POSTER PAPER

ABSTRACT

In this paper a set of visualization tools and related ideas are introduced for extracting and analyzing social network features present within persistent, multi-speaker, multi-topic, quasi-synchronous computer-mediated communication systems enacted over the internet, i.e., internet chatrooms. Preliminary models of these tools are applied to a real-world chatroom dataset. Results suggest the utility of such tools for enhancing the usability of digital text communications and the understanding of social structures and dynamics within the virtual world. Potentially promising visualization methods and areas of future research are discussed.

KEYWORDS: Information Visualization; Text Mining; Social Network Analysis; Visual Interface Design; Human-Computer Interaction; Computer-mediated Communication.

1. INTRODUCTION

Chat is a unique form of computer-mediated communication (CMC) that involves the exchange of digital text messages between one or more users online. It is unique in that it is persistent (logs are often accessible over time), can be multi-speaker, is often multi-topic, and is quasi-synchronous (communicators are often, but not required, to be on-line for meaningful communications to occur). These features characterize most internet chatrooms. There has been increasing interest in visualizing aspects of chat to improve users' experience, since this type of digital communication is missing key aspects of normal face-to-face communications (interactional coherence, social cueing and turn-taking,

etc.) or have unique features altogether that might be useful -- or perhaps detrimental -- to users (persistence of logs, can be quasi-synchronous, etc.) [1-3].

There has been particular interest in extracting and/or visualizing social network and CMC data (comprehensive reviews of digital text visualizations can be found in [4,5]). The purposes of analyzing/visualizing this information are varied, and range from serving as statistical benchmarks, for increasing social consciousness and interaction, for improving educational interactions, and for improving usability, navigability, and understanding of conversations and topics under discussion [6-9]. Mutton [10] notes the importance of visualizing or otherwise graphically representing social networks since this allows viewers "to determine facts about nodes and relationships between nodes more rapidly than examining the raw mathematical model" or the raw text data.

The present work is meant to discuss our preliminary work towards extracting social network information from chat text data, and propose some methods of visualizing this information. Our approach to extracting social network information primarily includes:

- analyzing messaging response time patterns of and between members (temporal proximity)
- determining message similarities (based on keywords of message content)

The point of analyzing these two pieces of information is for constructing networks of who people are communicating with, and who is in a group or a given conversation/topic at any given time. Our method is unique in that most other methods appear to construct networks simply based on who is a member of a self-selected and clearly delineated conversation or group, or who is responding explicitly to whom (via signaling such

as “respond to” functionality present in emails, blogs, or newsgroups). Thus, determining social structures in chat rooms is less straightforward than in other seemingly similar CMC domains due to the lack of explicit signaling, because in other domains these relationships are usually made obvious or are otherwise explicit [11].

The idea of using a “temporal proximity” approach (with other similar interactional measures) to infer social networks in chat was proposed by Mutton [10], who then visualized them using edge-and-node diagrams using IRC data. Our analysis adds the analysis of message content to infer message similarity, with the goal of making more accurate inferences regarding who belongs in any given network, and their relationships to others. Additionally, we discuss and present some alternative visualizations of this data beyond the traditional edge-and-node network diagrams.

The methods discussed herein are applied to a subset of real-world chat dataset (i.e., communications logs). Our methods were designed primarily for passive outside observers of persistent chat systems but could easily find utility with users for active communications and collaborations.

2. CHAT DATA COLLECTION AND ANALYSIS

We present next our methods of analysis that were applied to portions of a free public chat dataset that contains over 14 million messages. This chatroom is primarily for fans of music (it’s called “MusicBrainz”) and has kept a persistent text record of all chats for the last seven years and, as of this writing, the chatroom continues to the present day [12]. For the data plotted in Figure 1, all messages were analyzed, while most of the visualizations were developed using snippets of the dataset, usually several thousand lines long. All screen names were changed to protect anonymity.

The chat data was gathered though either an application programming interface (API) or through a web crawler. We stored the full text of each message, which included such content as the username of the person who posted it, the date and time it was posted, and the text message itself sent by the user. We computed additional metrics including a count for how many times each word was used in a message, times between sequential message postings, and a unique ID for the message. System messages were generally disregarded or deleted (such as “Tom has logged on”).

2.1. Keyword Similarity

Messages were assigned a similarity score based on the number of shared keywords they contained in common with other messages. In this context we consider keywords as words that are most informative about the substance of the conversation. Currently, the keywords in our tool are generated using a variant of the *term frequency - inverse document frequency* (TF-IDF) [13]. They are the words that occur the most often in the entirety of the user generated chat messages after the elimination of *stop-words* (common words that are primarily uninformative, including conjunctions, pronouns, prepositions, etc. [13]. Although, we intend to improve the keywords identification method by using standard stop-word resources (e.g., a language dictionary for word sense disambiguation as well as synonyms clustering), the current method was reasonably effective for our initial development purposes.

2.2. Temporal Proximity

Virtually all forms of human communication show a similar temporal pattern: most people respond quickly most of the time, but very rarely, a few people take a very long time to make a reply in a conversation, whether in an email conversation, a spoken conversation, a pen-pal letter, or internet chat. This typically rapid response rate of course makes human communication very efficient; if someone doesn’t respond within a reasonable amount of time, they either didn’t get the message, or are signaling something else (like “leave me alone” or perhaps “I’m too busy to respond right now”).

In regards to the present topic of extracting social network data, large time gaps between sequential messaging provide a potential hint that a conversation has ended (and a new one started), while small time gaps between messages suggest the possibility that communicators are directly engaging with each other. Presented in Figure 1 is

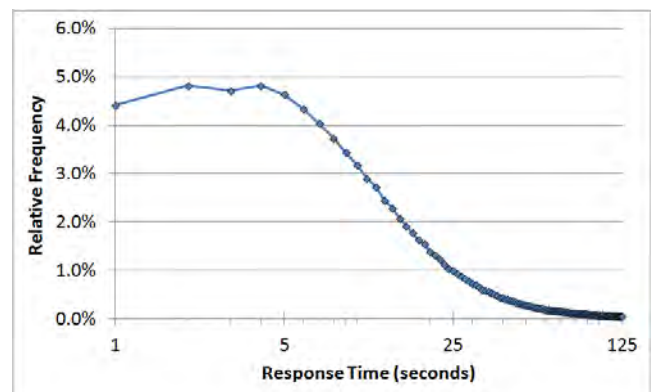


Figure 1. Relative Frequencies of Message Response Times (in seconds) in an Internet Chat Sample. Note the y-axis is a logarithmic scale.

the relative frequency distribution of the response times between sequential messages of the MusicBrainz chat data, consisting of over 14 million messages. This data was used to calculate the probability that a given message was “in response” to the preceding message; thus helping us infer who was talking to whom.

2.2. Other Content Analysis – Direct Addressing

Our current system employs some of the popular methods of identifying direct messages in chatrooms, i.e., if a chat message referenced an individual communicator by screen name, called “direct addressing” by Mutton (2004). This type of observation is common in digital text communications, and is provided by speakers to enhance the clarity of who the message is intended for (and who it might not be intended for). When using this method, limitations arise when the chat message does not contain the screen name of any user (there simply is no direct addressing), or it contains a variation of the intended individual’s screen name (either intentionally or by typographical error).

5. PRELIMINARY RESULTS AND DISCUSSION

The results of the preliminary analysis of our tool suggests that it functions fairly well in constructing social networks in chatroom text, which can then be further analyzed and/or visualized possibly based on some of the ideas presented in the next section.

One of the major problems we encountered was that extremely short, terse messages that contained few non-stop-words, particularly if given after a substantial delay to the previous message, resulted in fragmenting of the conversational thread at some inappropriate points, thus splitting up some networks that should not have been fractured. An example of this problem was the short reply “Yes” or “No, I don’t” after a substantial delay to a particular question. This issue might be alleviated in future work by giving greater weightings to common “reply” words like “yes,” “no,” “maybe,” “ok”, etc. that follow a message with a question mark; or perhaps by giving users the ability in the software to “point” or otherwise easily indicate which message is being referred or replied to, although this last remedy has its own unique set of problems related to usability and compliance.

6. SOME PROPOSED VISUALIZATION METHODS

Next, we present some of our social network visualization ideas. One particularly promising technique was simply the plot of what we called the “conversation cycles” for a chosen individual (as shown in Figure 2 for “Mike”). One temporal “cycle” indicates the individuals who communicated with the chosen person, and how often, between the chosen person’s communications in time. One can quickly assess for Mike’s conversations who are the active communicators with him, whether the groups seemed to be dyads, triads, or larger, and how often they spoke before Mike communicated next. The cycles are arranged vertically in time. Vertical patterns tend to indicate continuing/on-going conversations with individuals (for instance, with Jane for cycles 1 to 7, then with Lindsey for cycles 7 to 17), while horizontal patterns indicate groupings (in cycle 17, Mike, Jane, Fred, Lindsey and Joe all seemed to be interacting). This visualization method may suffer from scaling issues, but we thought it a unique way to try to infer groupings, see temporal patterns, and see individual interactions all within a single visualization.

Another visualization idea we considered is similar to a covariance matrix (but using relative frequency of contacts instead of co-variances), as presented in Figure 3. As can be observed, individual communicators and their connectedness to others can be quickly viewed. One disadvantage of this method, again, is that for very large numbers of communicators, these graphics would get complicated very quickly. Additionally, this shows connectedness between individual communicators, but does not indicate who is a member of which group (or network), the way an edge-and-node diagram easily does. For instance, while Jane and Mike talked frequently, did their conversations also include Bob? In other words, is Bob part of Jane’s and Mike’s social network...it is not immediately clear using the visualization technique in Figure 3, but it is a question a user or a viewer could conceivably wish to know the answer to.

6. CONCLUSIONS AND FUTURE WORK

Future research might assess scaling for these proposed visualizations; in other words, at what point (i.e., how many communicators does it take) for these types of graphics to become perceptually or cognitively unwieldy? What other visualization methods might be more useful? Are nodes-and-edges sufficient? How do the tasks of the user shape the required visualizations?

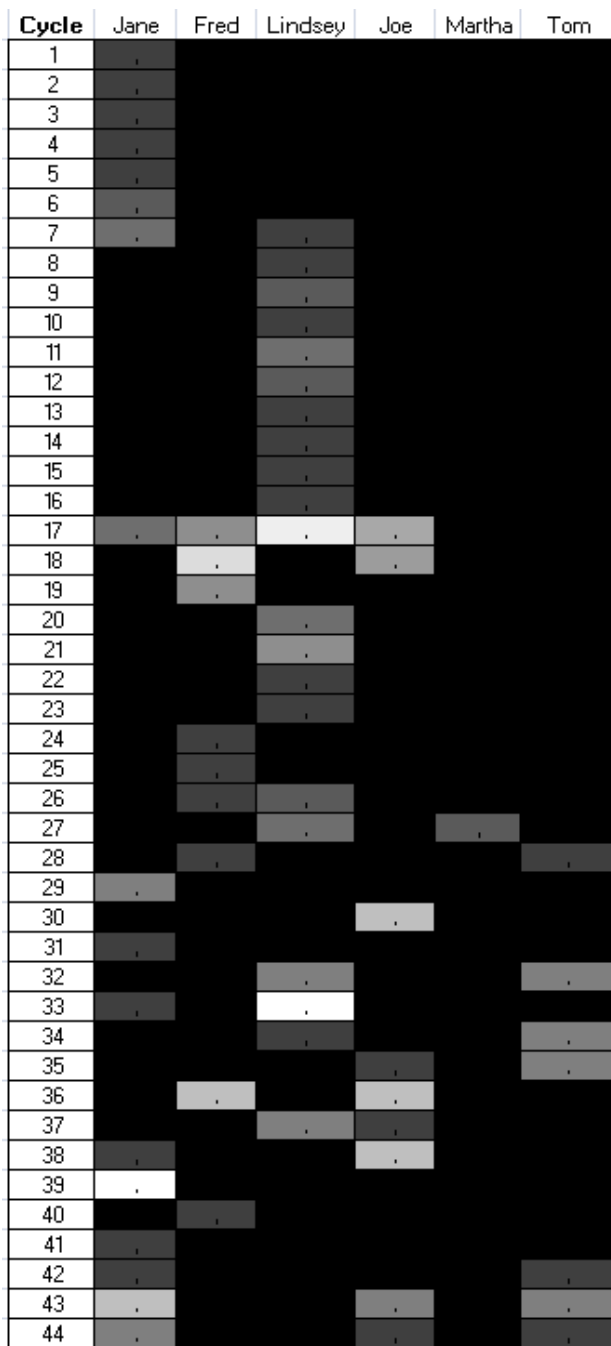


Figure 2. An example Conversation Cycle graphic for Mike. We can see that Mike spoke in individual conversations with Jane, then Lindsey, then interacted more sporadically with others. Horizontal patterns suggest social network groupings (as in cyce 17).

In this work, we attempted to build upon the similar work of others, particularly that by Mutton [10], to develop tools and visualizations that might aid in social network analysis of chat data. Our focus here was on the persistent logs of chatrooms, as much of the work has focused on extracting social data for other purposes than network

analysis, such as for enhancing the educational or social experiences of users. But these ideas could conceivably be applied to other types of distributed electronic text communications to aid in analysis or enhancing usability, such as Twitter or texting logs, IM, web-logs or “blogs”, newsgroup comment logs, e-mail, etc.

In the future, it was suggested that sentiment analysis could also be implemented to try to determine the emotional content of entries (such as the display of positive, negative, or neutral mood by message posters). Using such methods might be akin to “emotional” threading as opposed to conversational threading, allowing users to avoid highly negative conversations or single negative messages within conversations, for instance, or for users more interested in the substantive informational content in a thread rather than the emotionally-charged content, whether positive or negative.

We also would like to attempt the correlation of information from chat messages with information from other digital sources in order to establish relationships or trends. For example, if a number of the musicians being talked about in the chatroom have become successful over the past seven years (and hence information in other media e.g. newspapers). Or, perhaps it would be useful to study the rate/flow of information between various communication systems (how does info flow through other similar or even dissimilar networks, and why?).

For our own future work, we would like to continue in the development of these visualization tools. At this point our ideas are still at a preliminary stage of development and require much further technical work. Also, usability testing, particularly formal experimental techniques are needed. We hope the discussion and preliminary results presented in this work can help guide our future work and (hopefully) others working on similar technical problems in information visualization, text-mining, social network analysis, and distributed collaborative communication systems and technologies.

ACKNOWLEDGEMENTS

We wish to thank the Tec^Edge Summer at the Edge (SATE) program for its invaluable support, particularly that provided by Dr. Rob Williams from SATE and from Erica Edelman at SATE / Wright State University and fellow SATE member. Thanks to the Air Force Research Laboratory’s 711th Human Performance Wing Information Visualization section for their support, particularly from mentor Eric Geiselman. Thanks also to Dr. Steven Rogers at the Air Force Research Laboratory and Adam Rogers at Wright State University for their

many helpful and insightful discussions and ideas. This work was conducted by the Information Visualization Team at Tec^Edge's Summer at the Edge student program, under the guidance of the Air Force Research Laboratory's 711th Human Performance Wing Information Visualization section, over 10 weeks in the summer of 2010.

REFERENCES

- [1] Vronay, D., Smith, M., & Drucker, S., "Streaming media interfaces for chat," Proceedings of UIST, Asheville, NC, pp. 19-26, 1999.
- [2] Garcia, A.C., & Jacobs, J., "The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication," *Research on Language and Social Interaction*, vol. 32, issue 4, pp. 337-367, 1999.
- [3] O'Neill, J., Martin, D., "Text chat in action," Proceedings of ACM SIGGROUP conf. on Supporting Group Work, 2003.
- [4] Viegas, F., "Revealing individual and collective pasts: Visualizations of online social archives," Ph.D. dissertation in Media Arts and Sciences at the Massachusetts Institute of Technology, 2005. Available online at: http://smg.media.mit.edu/papers/Viegas/viegas_thesis.pdf.
- [5] Silic, A., "Visualizatin of text streams: A survey," Proceedings of the 14th International Conference on Knowledge-based and Intelligent Information and Engineering Systems: Part II, 2010.
- [6] Donath, J., Karahalios, K., & Viegas, F., "Visualizing conversation," in Proceedings of the 32nd Hawaii International Conference on System Sciences, 1999.
- [7] Tat, A., & Carpendale, S., "CrystalChat: Visualizing personal chat history," Proceedings of the 39th Hawaii International Conference on System Sciences, 2006.
- [8] Gibbs, W.J., Olexa, V., & Bernas, R.S., "A visualization tool for managing and studying online communications," *Educational Technology & Society*, vol. 9, issue 3, pp. 232-243, 2006.
- [9] Neumann, P., Tat, A., Zuk, T., & Carpendale, S., "KeyStrokes: Personalizing typed text with visualization," Proceedings of the Eurographics/IEEE-VGTC Symposium on Visualization, 2007.
- [10] Mutton, P. "Inferring and visualizing social networks on Internet Relay Chat," in Proceedings of the Information Visualization conference (InfoViz'04), IEEE Computer Society, Washington, D.C., pp. 35-43, 2004.
- [11] Herring, S. "Interactional coherence in CMC," *Journal of Computer-Mediated Communication*, vol. 4, issue 4, 1999, available at: <http://jcmc.indiana.edu/vol4/issue4/herring.html>, accessed January 2011.
- [12] "MusicBrainz IRC," [website], accessed October 2010, available: <http://musicbrainz.org/doc/IRC>.
- [13] Karen, S.J., "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, number 1, pp. 11-21, 1972).

Jane	Mike	Bob	Eric	Mary	Shelly	Fred	Joe	Kate	Lindsey	Martha	Paul	George	
	100	27	32	0	23	28	23	23	24	0	0	21	Jane
100		30	27	32	24	48	24	51	51	25	23	22	Mike
27	30		26	0	0	21	0	0	0	0	21	0	Bob
32	27	26		0	21	0	0	0	0	0	0	0	Eric
0	32	0	0		0	26	0	0	23	0	0	0	Mary
23	24	0	21	0		22	0	21	22	0	0	0	Shelly
28	48	21	0	26	22		30	34	53	0	25	0	Fred
23	24	0	0	0	0	30		0	26	0	0	0	Joe
23	51	0	0	0	21	34	0		44	23	21	0	Kate
24	51	0	0	23	22	53	26	44		0	23	0	Lindsey
0	25	0	0	0	0	0	0	23	0		0	0	Martha
0	23	21	0	0	0	25	0	21	23	0		0	Paul
21	22	0	0	0	0	0	0	0	0	0	0		George

Figure 3. An Example Contact Matrix. This displays who is talking to whom, and with what relative frequency.

How much camera separation should be used for the capture and presentation of 3D stereoscopic imagery on binocular HMDs?

John McIntire, Paul Havig, Eric Geiselman, and Eric Heft
711th Human Performance Wing / RHCvZ, Air Force Research Laboratory
2255 H Street, Wright-Patterson AFB, OH 45433-7022

ABSTRACT

Designers, researchers, and users of binocular stereoscopic head or helmet-mounted displays (HMDs) face the tricky issue of what imagery to present in their particular displays, and how to do so effectively. Stereoscopic imagery must often be created in-house with a 3D graphics program or from within a 3D virtual environment, or stereoscopic photos/videos must be carefully captured, perhaps for relaying to an operator in a teleoperative system. In such situations, the question arises as to what camera separation (real or virtual) is appropriate or desirable for end-users and operators. We review some of the relevant literature regarding the question of stereo pair camera separation using desk-mounted or larger scale stereoscopic displays, and employ our findings to potential HMD applications, including command & control, teleoperation, information and scientific visualization, and entertainment.

Keywords: stereo pair, camera separation, stereoscopic photography, virtual environments, micro displays, human factors, stereographics, stereoscopy, baseline selection, orthostereopsis

1. INTRODUCTION

Users of three-dimensional (3D) stereoscopic displays are consistently faced with the tricky issue of what imagery to present via their particular display set-up. More often than not, stereoscopic imagery must be created in-house with a 3D graphics program or stereo photos must be carefully captured with one or more cameras. In such situations, the question arises as to what camera separation (real or virtual) is appropriate or desirable, often leading to a pain-staking process of trial-and-error. From an Air Force operational perspective, it's not immediately obvious what "rule-of-thumb" to use or what camera separation calculator works best, so operators might have no idea how far apart UAV or satellite sensors or teleoperative robotic cameras should be separated if three-dimensional imagery is requested or required. This is an important issue for any non-volumetric stereoscopic display systems (shutter, polarized, auto-stereos, etc.) and for head or helmet-mounted displays (HMDs). In this work, we will discuss some of the existing literature on camera separation for stereoscopic imagery and displays, including the Human Factors implications, and relate our findings to the use of binocular stereoscopic HMDs.

2. PREVIOUS WORK ON STEREO CAMERA SEPARATION

2.1 Rule-of-thumb ratios

In the depth perception, 3D photography, and optical engineering literature, there exists at least some guidance regarding the issue of camera separation (or *camera baseline*). Rules-of-thumb abound, but interestingly these vary widely in their suggestions. For instance, if one is capturing stereoscopic imagery in a computer-generated virtual environment, some suggest a separation ratio of 10-to-1 (i.e., for every 10 units in distance from the virtual scene, separate the two cameras by 1 unit).¹ The Polaroid Interocular Calculator, developed by Kittrosser in 1952,² assumed that a ratio of 24-to-1 was most desirable for real-world scenes, but there is not strong empirical support for these ratios.³ Other suggestions can be found for stereo photography ranging from a 60-to-1 ratio,^{4,5} a 30-to-1 ratio,⁴ or a 20-to-1 ratio,⁶ all with little or no empirical or Human Factors evidence in support of them. So which ratio should be used, if any?

2.2 The camera/scene space versus the viewer/display space

Jones et al.³ clearly point out the important distinction between the camera/scene space and the viewer/display space for stereoscopy, which explains the intertwined relationships between the viewer, the scene, the camera, and the display. Both spaces are important to consider in their own right to understand the transformations that occur between them (as shown in Figure 1).

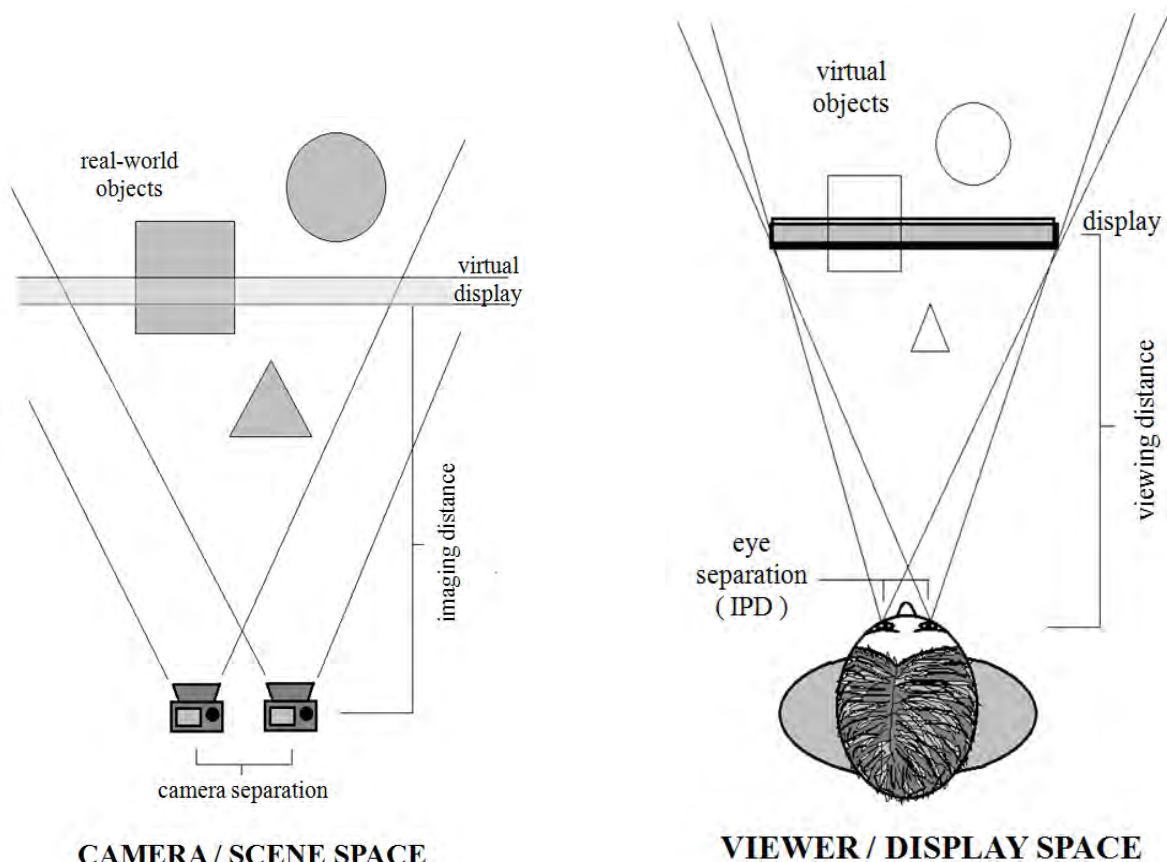


Figure 1. An illustration of the distinction between the camera/scene space parameters and the viewer/display space parameters, as suggested by Jones et al.³ This distinction is useful for understanding the relationships between the camera, the scene, the viewer, and the display in stereoscopy.

- The **camera/scene space** – factors involving the capture or generation of stereoscopic imagery (real or virtual), such as distance between the camera and the objects of interest in the scene, camera features (field-of-view [FOV], film width, lens focal length), imaging methods, etc.
- The **viewer/display space** – factors involving the presentation and viewing of stereoscopic imagery, such as distance between the viewer and the display, the display field of view, viewer eye separation (inter-pupillary distance or IPD), stereoscopic display method, display viewing volume, etc.

Once we carefully consider this distinction, we can easily see why simple rule-of-thumb ratios do not often work. While the ratios take into account some variation in the camera/scene space (by positioning the cameras more or less apart – according to a fixed ratio that depends on the estimated distance to the important objects in a scene), the ratios do not consider the ultimate presentation of the imagery or how it will be viewed by the viewer. So, unfortunately, rules-of-thumb are often not very helpful to the developer or user of stereoscopic imagery because trial-and-error is still often necessary to get natural, comfortable depth on any particular display. Jones et al.³ used this camera/scene versus viewer/display space distinction to aid in the creation of interocular calculators, a topic which will be discussed next.

2.3 Camera separation algorithms

To combat the problems with using rules-of-thumb ratios, several camera separation algorithms and/or methods have been developed by engineers and photographers over the years.³ Lipton calculated and published recommended camera separations for stereo photography.⁷ He created a series of tables that gave the recommended minimum and maximum camera separation, which depended upon the distance of the scene and the width of the film (i.e., film format). A

potentially serious problem with these tables is that they assume the use of a converging or “toed-in” camera set up, instead of aligning them in parallel. The use of convergent imagery for stereoscopic displays is today highly discouraged due to the distortions that are caused, particularly that of vertical disparity.^{8,9,10} Kittrosser’s Polaroid Interocular Calculator, mentioned previously, takes as input the scene parameters such as distance to the near and far point-of-interest in the scene and calculates the necessary camera separation assuming a particular ratio. This strategy takes account of *some* factors in the camera/scene space (like the region of interest in the scene and its depth range), but does not account for variations in the viewer/display space.

Williams and Parrish¹¹ gathered experimental human factors data on ranges of comfortable binocular fusion, and used these data to derive algorithms for determining virtual camera separation based on parameters from the two spaces discussed above. The algorithms map a depth range in the imaging space to a human factors specified depth range in the display/viewing space. The mappings are meant to ensure that the stereoscopic imagery does not result in un-fusable double imagery (diplopia) or other viewer discomfort. Depth scenes are scaled (either compressed or expanded) to fall within the experimentally-derived acceptable range from -25% to +60% of the viewer-to-display distance, according to their results.¹¹

This algorithm and other similar ones^{3,10} are likely to introduce depth distortions (compressions/expansions) which could conceivably alter the veridical perception of depth in a scene. This could be problematic in many conceivable situations, particularly those involving fine hand-eye coordination within the near depth field where accurate depth perception is absolutely critical (teleoperative systems like tele-surgery or tele-robotics). Although some human factors studies suggest that the lack of veridical depth representations – also referred to as *true-to-scale 3D* and *orthostereopsis*¹² – does not cause bizarre perceptual distortions as might be expected,¹³ other research suggests otherwise.⁸ However, in terms of performance, orthostereopsis does not seem to be necessary for performing depth tasks at an acceptable level of performance.^{14,15} This is probably because “people’s understanding of the global layout of objects in space does *not* come primarily from stereoscopic depth cues...kinetic depth and linear perspective cues are more important.”¹³ So non-veridical stereoscopic representations are unlikely to greatly alter depth task performance, as long as distortions are not too large, but it is important that this claim be convincingly verified (and the acceptable range established) since other Human Factors results suggest possible perceptual disturbances.

3. STEREO CAMERA SEPARATION FOR HMD’S

The interocular calculators in the literature generally are not directly applicable to HMDs,¹⁰ because they assume that the images intended for each eye spatially overlap at the display surface (as in most stereoscopic systems), but this is obviously not true for binocular HMDs, which have separate displays directly positioned in front of each eye (and are usually collimated or focused at optical infinity). This distinction is shown in Figure 2, and implies that somewhat different transformations (i.e., interocular calculators) are needed when presenting images stereoscopically via binocular HMDs.

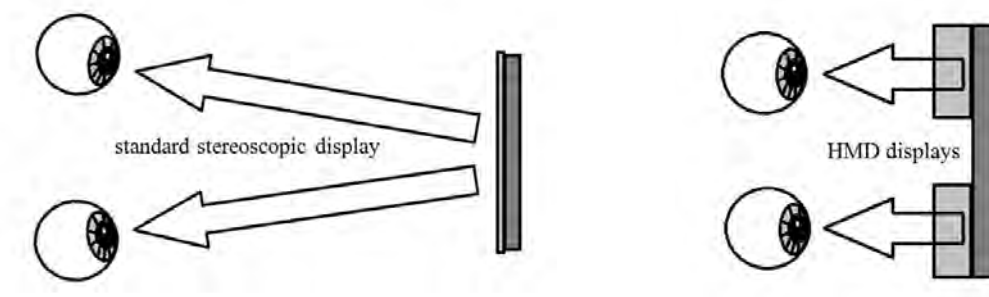


Figure 2. An illustration of the distinction between the stereoscopic display methods for traditional stereoscopic displays (which present a separate view to each eye from the same display surface) and binocular HMDs (which use two different displays, one for each eye).

Next, we present camera separation calculations intended for imaging scenes that will be displayed on HMDs. Luckily, the calculations are relatively straightforward in the case of HMDs because both the imaging space and the viewing space use parallel configurations. As mentioned earlier, using a parallel camera configuration is highly recommended in the literature due to the many types of imaging distortions caused by convergent or “toed-in” sensors. Since HMDs typically use a parallel presentation technique requiring no ocular convergence (as demonstrated in figure 2), this creates a linear relationship between objects’ depths in the imaging space and virtual objects’ depths in the display space.^{10,12}

In other words, since *image disparity* leads to *display disparity* which leads to *retinal disparity*, an orthostereoscopic depth representation can be presented to the user’s retinas by matching the image disparity to the display disparity and matching viewer eye separation to the camera separation. This is a relatively trivial problem when using both parallel cameras and parallel display configurations with matching FOVs because this does not require image clipping/cropping or off-setting the image sensors from each lenses’ optical axes, as other methods do.^{3,10,17} Thus, presentation of 3D scenes on HMDs can straightforwardly present orthostereoscopic cues if camera FOV and display FOV are equated. If these FOVs are not matched, size compression or expansion of the imagery will occur, but the imagery can still be manipulated post hoc (clipped or expanded) to be orthostereoscopic; in any case, the depth relations in the scene will advantageously remain linearly related. The following discussion of camera separation for orthostereoscopic display assumes that there is an acceptable real-world reference to size and so may or may not be applicable to virtual environment displays that have no real-world referent (e.g., highly abstract data sets).

3.1 Camera separation for an orthostereoscopic HMD display

As discussed, if an orthostereoscopic representation is desired, then the image FOV must be matched to the display FOV. The easiest way to accomplish this is to match the *camera FOV* to the display FOV, which automatically creates imagery with a FOV matching the display and requires no further image manipulation. The two cameras should then be separated precisely by the viewer’s IPD to present an orthostereoscopic display. Camera FOV can be calculated as follows, from the chosen dimension d (height/width of the imaging sensor) and the effective focal length f (note that further calculations may be needed for f , including the magnification factor and the stated lens focal length if used for macro photography):

$$\text{Camera FOV} = 2 \arctan \frac{d}{2f}$$

If it is undesired, impractical, or impossible to directly manipulate the camera FOV, but an orthostereoscopic representation is still desired, then the image FOV must be manipulated after capture. The ratio of the camera FOV to the display FOV gives the expansion/compression factor that is required for image manipulation resulting in an orthostereoscopic display:

$$\text{Image Expansion/Compression Ratio} = \frac{\text{Camera FOV}}{\text{HMDFOV}}$$

The benefit of post-hoc image manipulation is that the proper camera separation remains precisely the viewer’s inter-pupillary distance or IPD, since the imagery and thus the corresponding inter-camera disparities are scaled to match their orthogonal sizes.

3.2 Camera separation for a non-orthostereoscopic HMD display, and incorporating Human Factors results

Any strictly orthostereoscopic representation has the potential to display non-apparent depth, non-fusible depth, or uncomfortable depth to a viewer; the ultimate appearance depends on the original imaging scene. But by using a non-orthostereoscopic display, and manipulating camera separation, we can ensure that any stimuli presented to a viewer will fall within an acceptable and comfortable range of binocular fusion. Thus, the use of non-orthostereoscopic displays may in most cases be an inevitable consequence of stereoscopic imaging and display, as suggested by Jones et al.³ and Lipton.⁷ Depth range manipulations effectively compress/expand the depth relationships in the scene to ensure disparity is not over or under stimulating. The Human Factors work of Williams and Parrish¹¹ suggest that this acceptable range is about -25% / +60% of the distance to the display when using standard desktop stereoscopic displays.

But is an HMD with collimated displays limited to this same range? On an HMD, just as on any other display, any uncrossed (divergent) disparities greater than ½ the viewer’s IPD will require ocular divergence beyond parallel lines of

sight and hence be uncomfortable and/or un-fusable, which sets a clearly definable limit. But this should not be problematic with parallel camera and parallel display configurations because objects imaged at infinity will appear along parallel lines-of-sight on the displays, producing no interocular disparity. Optical infinity for human vision is for practical purposes set by optometrists at about 20 feet. This sets the far-point range for our display space at 20 feet (~5716 mm). The near point distance, according to Williams' and Parrish's assumption of -25%/+60%, then, would be 9.375 feet, which we'll round to 10 feet (~2858 mm), according to the following calculations:

Williams & Parrish recommendations for a comfortable depth volume in the viewer/display space:

$$\text{Near Point} = -25\% \text{ distance to screen} = 0.75 \times d$$

$$\text{Far Point} = +60\% \text{ distance to screen} = 1.60 \times d$$

To apply these assumptions to HMDs (remember, there is not strictly a "display plane distance" in the HMD), we just want to derive the near and far points from Williams' and Parrish's work:

$$\text{Far Point} = \text{near optical infinity} \sim 20 \text{ feet}$$

$$\text{Far Point} = 1.60 \times d, \text{ therefore } d = 12.5 \text{ feet}$$

$$\text{Near Point} = 0.75 \times d = 0.75 \times 12.5 \text{ feet} = 9.375 \text{ feet} \sim 10 \text{ feet} \sim 2858 \text{ mm}$$

So we want to make sure that any image we take in the camera/image space will only contain depths (disparities) ranging from ~10 feet to optical infinity. In order for this to work, the following ratios must hold when transforming from the near point of the camera/scene space to the near point of the viewer/display space:

$$\frac{\text{Camera Separation}}{\text{Near Point (C/S space)}} = \frac{\text{IPD}}{\text{Near Point (V/D space)}}$$

$$\text{Camera Separation} = \frac{\text{IPD (in mm)}}{2858 \text{ mm}} \times \text{Near Point (C/S space)}$$

If camera FOV and HMD FOV are matched, then the above equation is all that is needed to determine the appropriate camera separation (in mm), after determining the viewer IPD (in mm) and the distance to the near point of interest in the camera/scene space (again, in mm).

We can get an idea of where some of the rules-of-thumb ratios discussed earlier might have come from: notice that the Near Point in the Viewer/Display space (2858 mm) is divided by the IPD, and since the typical IPD is ~63 mm, then this would produce a ratio of display distance-to-IPD of about 45. This ratio could then be retained in the camera/scene space, by separating the cameras 1 unit for every 45 units of distance to the near point-of-interest in the scene. Interestingly enough, this ratio (45) falls somewhere in the middle of the many suggested rule-of-thumb ratios mentioned earlier. Assuming that the camera FOV is roughly equivalent to the display FOV, this ratio would probably provide nice, comfortable, fusible depth scenes on an HMD.

If camera FOV and HMD FOV are not equivalent, problems can quickly emerge in terms of distortion (compression/expansion) and too little/too much depth. Next, we modify the above formula to take account of the ratio between camera FOV and HMD FOV.

$$\text{Camera Separation} = \frac{\text{IPD (in mm)}}{\text{Near Point (V/D space)}} \times \text{Near Point (C/S space)} \times \frac{\text{Camera FOV}}{\text{HMDFOV}}$$

Therefore, these calculations suggest that in the viewer/HMD space, we should provide a display disparity range corresponding to distances between 10 feet (near point) and 20 feet (far point – near optical infinity). Doing so should provide a very comfortable depth volume for viewers of virtually any scene, and regardless of differences of camera FOV and display FOV. Additionally, we see that if Cam FOV = HMD FOV, that ratio computes to 1, and is effectively equivalent to the preceding simpler formula. This calculation also allows for adjustment of the near point distance in the virtual/display space, so that other distances that differ from Williams' and Parrish's range might be used, and could be updated based on personal preference or further human factors results.

4. DISCUSSION, CONCLUSIONS, AND FUTURE WORK

The short answer to the question “how much camera separation should be used if displaying 3D imagery via binocular HMDs?” is: *it depends*, primarily on the task at hand. For very important or critical tasks involving intricate hand-eye coordination in the near-field, like remote teleoperations of bomb disposal robot arms, tele-surgery, etc., then a strictly orthostereoscopic depth representation is at this time recommended (formulas given in section 3.1), simply because it is not completely clear what the perceptual and performance effects might be under non-orthoscopic conditions. Further work on this issue is important. If strict orthostereopsis is not deemed critical, then the formulas in section 3.2 can provide camera separations for expansion/compressions of imaged scenes, so that the displayed depth volumes fall within natural, comfortable viewing ranges.

Problems with showing depths on any stereoscopic displays (HMDs or not) are exacerbated by the unavoidable decoupling of eye accommodation (focus) and ocular vergence angle, which are normally coupled in everyday life. This natural coupling explains why humans do not find large ranges of objects’ depths to be strange or annoying in the real world, and perhaps why diplopia (double imagery) goes almost completely unnoticed, since the diplopic images are very much out of focus anyway. But this is not true for HMDs, or any current non-volumetric stereoscopic display system, because if a viewer’s eyes are accommodated to the display surface, then the entire display surface is in focus, even objects intended to appear at difference depths. Volumes have been written about this accommodation-vergence mismatch problem and its implications for 3D displays. We only touch on it to point out that simple geometric display considerations may not tell the whole story, and that Human Factors issues suggest that the displayed depth range may need to be smaller than one might initially expect in order to be comfortable and usable. Further work is recommended on this topic.

We also suggest future work on the perceptual and performance implications of using unnaturally large or small camera separations. For instance, *hyperstereoscopic* displays (also “hyperstereo” or “telestereo”) make use of patently unrealistic camera separations, far beyond the average human IPD of 63 mm, and often in conjunction with lens magnification. The opposite of hyperstereo is *hypostereopsis* or *microstereopsis* where the virtual IPD is unnaturally small. These camera separation baselines could easily induce distortions like the *puppet-theater effect* (objects look unnaturally small and tiny) or the *cardboard effect* (objects look like flat paper cut-outs, although depth relationships appear maintained),¹⁶ or result in either diplopic imagery or absence of stereoscopic depth. While it is clear that perceptual distortions can and do occur from these types of manipulations, little human factors work has evaluated the effects on depth task performance, and those that do provide somewhat mixed results.^{8,14,15,17} So further work is recommended on this topic, so that we can know for sure whether orthostereoscopic representations are necessary for critical tasks like tele-surgery, teleoperative bomb disposal, etc. (as discussed in section 2.3).

Some experimental results suggest little impairment to depth performance if smaller IPDs are used.^{14,15} This is important to verify, and if true, could potentially alleviate many human factors issues related to oculomotor discomfort when using HMDs, or stereoscopic displays in general, because small amounts of disparity could be used without negatively impacting performance. Likewise, the perceptual and performance implications of using hyperstereoscopic systems, and their possible relationships with simulator sickness symptoms and display motion versus self-motion, etc. will be very important to study. Interest in remote teleoperation of dynamic vehicles and machinery is growing fast, as well as interest in using and exploring totally virtual worlds. It is our hope to progress the research on this topic so that viewers and users can use these amazing new technologies comfortably, safely, and effectively.

REFERENCES

1. “Stereo Rendering.” http://graphics.stanford.edu/~cek/rayshade/doc/guide/section2_5_4.html
2. Kittrosser, S. (1998). “Photography in the service of stereoscopy,” *Journal of Imaging Science and Technology*, 42, pp. 295-300, July/August.
3. Jones, G., Lee, D., Holliman, N., Ezra, D. (2001). “Controlling perceived depth in stereoscopic images.” *Proceedings of SPIE*, 4297, 42-53.

4. Lee, D. (2009). "Lens Separation in Stereo Photography."
http://www.berezin.com/3d/Tech/lens_separation_in_stereo_photog.htm
5. Wikipedia, "Stereoscopy: Imaging Methods," accessed 4 Jan 2011 from:
http://en.wikipedia.org/wiki/Stereoscopy#Imaging_methods
6. Bourke, P. (1999). "Stereo Pair Photography (The Low Cost Way)."
<http://local.wasp.uwa.edu.au/~pbourke/miscellaneous/stereophoto/>
7. Lipton, L. (1982). *Foundations of the Stereoscopic Cinema: A Study in Depth*, Van Nostrand Reinhold Company.
8. IJsselsteijn, W. A., de Ridder, H., & Vliegen, J. (2000). "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration." *IEEE Transactions on Circuits and Systems for Video Technology*, 10(2), 225-233.
9. Grinberg, V. S., Podnar, G., & Siegel, M. W. (1994). "Geometry of binocular imaging." *Proceedings of the IST/SPIE Symposium on Electronic Imaging, Stereoscopic Displays and Applications*, 2177.
10. Woods, A., Docherty, T., & Koch, R. (1993). Image distortions in stereoscopic video systems. *Proceedings of the SPIE, Stereoscopic Displays and Applications*, vol. 1915.
11. Williams, S. P., & Parrish, R. V. (1990). "New computational control techniques and increased understanding for stereo 3-D displays." *Proceedings of SPIE*, vol. 1256, pp. 73-82.
12. Diner, D. (1991). "A new definition of orthostereopsis for 3-D television," *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1053-1058, October.
13. Ware, C., Gobrecht, C., & Paton, M. A. (1998). "Dynamic adjustment of stereo display parameters," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, vol. 28, issue 1, pp. 56-65.
14. Ellis, S. R., Fishman, J. M., Hasser, C. J., & Stern, J. D. (2005). "Effect of reduced stereoscopic camera separation on ring placement with a surgical telerobot," *Proceedings of SPIE*, vol. 5664, pp. 372-379.
15. Rosenberg, L. B. (1993). "The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks," *Proceedings of IEEE VRAIS '93*, pp. 27-33.
16. Yamanoue, H., Okui, M., Yuyama, I. (2000). "A study on the relationship between shooting conditions and cardboard effect of stereoscopic images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, issue 3, pp. 411-416.
17. U.S. Army Aeromedical Research Laboratory (2009). *Helmet-Mounted Displays: Sensation, Perception, and Cognition Issues*. Eds. Rash, C. E., Russo, M. B., Letowski, T. R., & Schmeisser, E. T.

Visualizing Weighted Networks: A Performance Comparison of Adjacency Matrices versus Node-link Diagrams

John P. McIntire^{a*}, O. Isaac Osesina^b, Cecilia Bartley^b
M. Eduard Tudoreanu^b, Paul R. Havig^a, Eric E. Geiselman^a

^a 711th Human Performance Wing / RHCW, US Air Force Research Laboratory, Dayton, OH, USA;

^b Dept. of Information Science, University of Arkansas at Little Rock, AR 72204, USA

ABSTRACT

Ensuring the proper and effective ways to visualize network data is important for many areas of academia, applied sciences, the military, and the public. Fields such as social network analysis, genetics, biochemistry, intelligence, cybersecurity, neural network modeling, transit systems, communications, etc. often deal with large, complex network datasets that can be difficult to interact with, study, and use. There have been surprisingly few human factors performance studies on the relative effectiveness of different graph drawings or network diagram techniques to convey information to a viewer. This is particularly true for weighted networks which include the strength of connections between nodes, not just information about which nodes are linked to other nodes. We describe a human factors study in which participants performed four separate network analysis tasks (finding a direct link between given nodes, finding an interconnected node between given nodes, estimating link strengths, and estimating the most densely interconnected nodes) on two different network visualizations: an adjacency matrix with a heat-map versus a node-link diagram. The results should help shed light on effective methods of visualizing network data for some representative analysis tasks, with the ultimate goal of improving usability and performance for viewers of network data displays.

Keywords: visualization, visual analytics, human factors, adjacency matrix, heat maps, node-link diagrams, sociograms

1. INTRODUCTION

Fields such as social network analysis, genetics, biochemistry, intelligence, cybersecurity, neural network modeling, transit systems and communications often deal with large, complex network datasets that can be difficult to interact with, study, and use. Although there is growing research into developing and testing visualization methods for such large and high-dimensional datasets, there are surprisingly few empirical performance (human factors) studies addressing which ones work best and why. With regard to network data, there are few studies which have investigated the relative effectiveness of different graph drawings or network diagram techniques to convey information to an observer [1,2]. Such knowledge about the differential effectiveness of network data visualizations can be important for many areas of academia, applied sciences, the military, and the public. This study compares the effectiveness of two network visualization methods (an adjacency matrix with a heat-map and a node-link diagram) for portraying this type of information. The ultimate goal is to improve visualization usability for human decision makers.

1.1 Related Work

A series of experiments by Purchase and colleagues [3-6] suggested that node-link diagrams should minimize link crossings and avoid using bends, kinks, or turns in the links, in order to improve graph reading performance; otherwise, it makes little difference for graph readability how a node-link diagram is portrayed. They found little relation between graph aesthetics and performance (in terms of ‘understanding’ a graph). They also found (as have many others) that node-link diagrams simply do not handle well the problem of *scaling*: as the underlying networks get bigger (above 20 to 30 nodes), the graphs are more difficult (or sometimes impossible) to understand and/or use. For “traditional” node-link network diagrams, the scaling problem is almost certainly due to occlusion of the nodes/links by the large numbers of other nodes/links. This scaling issue in particular makes finding alternative network visualization methods important in order to handle the increasing sizes of available data. A compelling possibility is the display of connectivity or adjacency matrices, potentially augmented with other visualization techniques, like heat maps, to portray continuous values (for weighted network data).

In a study similar to ours, Ghoniem, Fekete, & Castagliola [7] examined user performance on node-link versus matrix-based network visualizations. They tested performance on seven generic graph-based tasks while manipulating the size and density (connectivity) of the networks. For networks larger than about 20 nodes, they found matrix-based visualizations were better for most tasks (the one exception being a path-finding task, such as ‘find the shortest path between nodes A and B,’ which was better when using node-link diagrams). Their results suggest that matrix-based visualizations should work well for large network datasets on a variety of network tasks, and their use should be encouraged despite observers’ relative unfamiliarity with this visualization method. Our research project attempts to replicate several key aspects of this study while additionally studying features intended to account for weighted network data: (1) the use of heat maps (as opposed to binary displays) in the matrices and, (2) luminance/color-varying links in the network diagrams.

2. METHODS

2.1 Equipment & Stimuli

Two possible visual stimuli were used: a node-link network diagram and a connectivity (or adjacency) matrix with a heat map (color/intensity is redundantly mapped to the link weights, with lighter colors indicating stronger links). The node-link network diagram consisted of 30 numbered nodes connected by weighted links/edges (again, with color/intensity being mapped to the link weights), as shown in Figure 1 (left). Following Purchase et al.’s graph aesthetics findings [3-6] (layout methods had little effect on performance as long as link bendings and/or crossings were minimized), no specific graph layout algorithm was used for this visualization. After multiple randomized spatial iterations the layout that minimized clutter/overlap of the links and nodes was selected. The adjacency matrix heat map displayed the same network data as the node-link diagram, and used the same color-coding scheme (to represent link weights), albeit in a different visual form. Row and column numbers indicated corresponding nodes, and were arranged in numerical order along both axes, thus the matrix was symmetrical due to this arrangement as shown in Figure 1 (right). Stimuli were presented via a laptop computer with a 15-inch LCD widescreen monitor (resolution of 1280H x 800V). A stopwatch was used by the experimenter to record task completion times, and participants recorded their responses using paper forms and a pen/pencil.

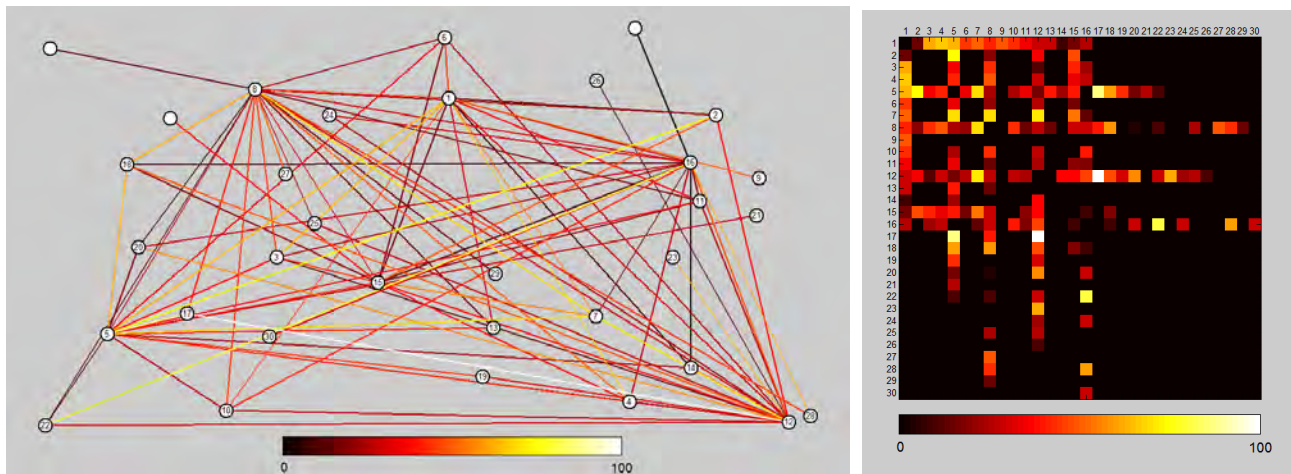


Figure 1. *Left:* The node-link network diagram visualization. *Right:* The adjacency matrix heatmap visualization. Both visualizations represent the same underlying network data. Connection (link) strength is represented by color/intensity (redundantly coded), with corresponding legends shown below both visualizations.

2.2 Participants

This experimental protocol was reviewed and approved as posing minimal risk for human participation by the Wright-Patterson Air Force Base Institutional Review Board (IRB). A total number of 20 volunteers participated; 14 males and 6 females. The inclusion criteria were normal or corrected-to-normal visual acuity (self-reported as having had a professional eye examination within the last 12 months). Participants were volunteers recruited from active duty military and DoD civilians and contractors from the 711th Human Performance Wing at Wright-Patterson Air Force Base.

Recruitment was conducted by word of mouth and/or email inquiries. Potential volunteers were asked if they would like to participate in a brief study on network visualizations; a short verbal description was provided and a copy of the Informed Consent Document was available for perusal. There was no restriction based on age or gender, nor any need to use these as limiting criteria. Total testing time for each participant lasted no more than 20 minutes, although there were no formal time limits for completion of the tasks. No compensation specific to this activity was provided to participants.

2.3 Experimental Procedure

Participants were randomly assigned to either the matrix visualization or the node-stick diagram stimulus for completion of tests in the first three task categories (link-finding, link strength estimation, and node connectivity estimation) and the opposite visualization for testing in the fourth category (node-finding). This resulted in a between-subjects experimental design where each participant was exposed to both visualizations, but they performed different tasks with each. Task categories were performed in the order presented below.

2.4 Link-finding Task

Participants were instructed to determine as quickly and as accurately as possible whether two defined nodes had a direct link/connection between them. There were 15 test items for this task.

2.5 Link Strength Estimation Task

If a link was found between the target nodes, participants were asked to indicate the apparent strength of the connection based upon the color-coding scheme (described in section 2.1.), ranging in possible values from 0 to 100.

2.6 Node Connectivity Estimation Task

This task was not timed. Participants were instructed to indicate the five nodes that were the most densely interconnected with the other nodes (had the most links directly connected to it). Participants were instructed to guess the most connected nodes by visual estimation instead of counting the links by hand.

2.7 Node-finding Task

For this task, the visualization type was changed from the previous tasks. Participants were again timed and instructed to respond as quickly and as accurately as possible whether two given nodes share a common neighboring node, and to indicate the number of that shared node (if there was one).

3. RESULTS AND DISCUSSION

All of the following significance tests were conducted using two-tailed Welch's *t*-tests assuming unequal variances, with a significance level of $\alpha = 0.05$.

3.1 Link-finding Task

On the link-finding task (determining whether two nodes were directly connected), the matrix visualization resulted in significantly faster task completion times (faster by 56%) than the node-link visualization ($t=4.90$, $p=.001$). The matrix visualization also resulted in less variable task completion times (standard deviation of 34 s versus 92 s for the node-link visualization). Accuracy was comparable across both visualizations (96% for the matrix, 94% for the node-link; $t=0.37$, $p=.721$), but the node-link exhibited greater variability in accuracy (standard deviation of 12% versus 4% for the matrix visualization). These results are shown in Figure 2.

3.2 Link Strength Estimation Task

On the link strength estimation task, judgments were comparable in accuracy across the visualization types (± 6.74 for the matrix, ± 6.19 for the network; judgments of strength ranged anywhere on a scale from 0 to 100; $t=0.71$, $p=.494$). Variability of judgment accuracy, as measured by standard deviation, was slightly larger for the node-link visualization at 1.84 versus 1.34 for the matrix visualization.

3.3 Node Connectivity Estimation Task

For judgments of which nodes were the most interconnected, the difference between the two visualization types in terms of accuracy were non-significant ($t=1.38$, $p=.197$) but the trend appeared to favor the matrix visualization (93%) versus

the network visualization (87%). Variability in accuracy was comparable across the visualizations (9.6 for the matrix versus 9.4 for the node-link visualization, in units of standard deviation).

3.4 Node-finding Task

For the task of identifying a commonly shared node between two given nodes, the matrix visualization resulted in significantly faster completion times (faster by 45%) than the node-link visualization ($t=3.03, p=.013$). Variability in task completion times was just a bit higher when using the matrix visualization compared to the node-link visualization (99 versus 92 units of standard deviation, respectively). Accuracy appeared slightly better with the node-link visualization (95%) versus the matrix visualization (86%) but this difference was non-significant ($t=1.00, p=.340$). However, variability in accuracy was noticeably higher for the matrix visualization at 27 units of standard deviation versus the node-link’s 7 units. Results for the node-finding task are shown below in Figure 3.

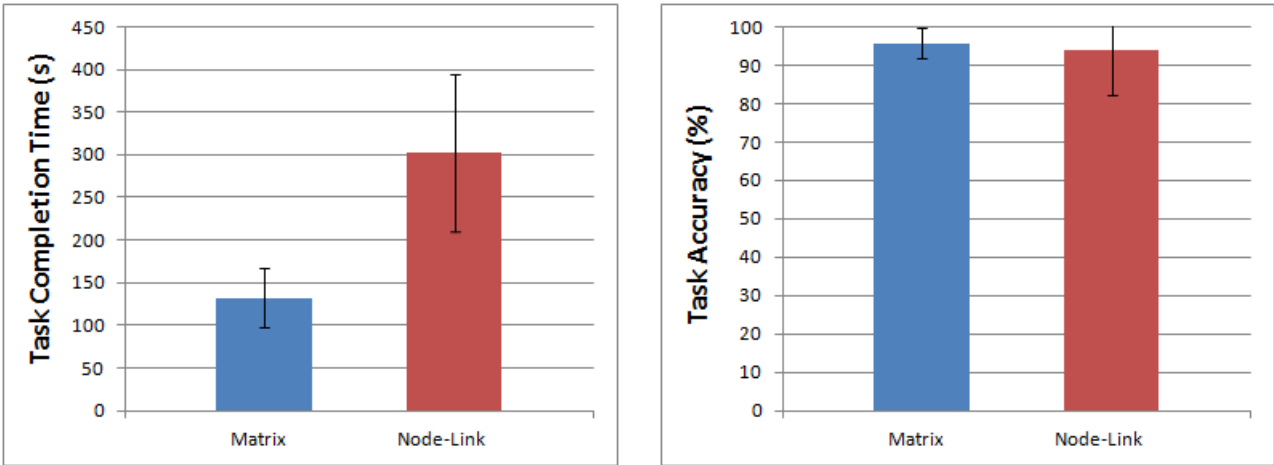


Figure 2. *Left*: Total task completion times (in seconds) on the link-finding task for the matrix visualization and the node-link visualization. Error bars represent +/- one standard deviation from the mean. *Right*: Task accuracy on the link-finding task. Error bars again represent +/- one standard deviation.

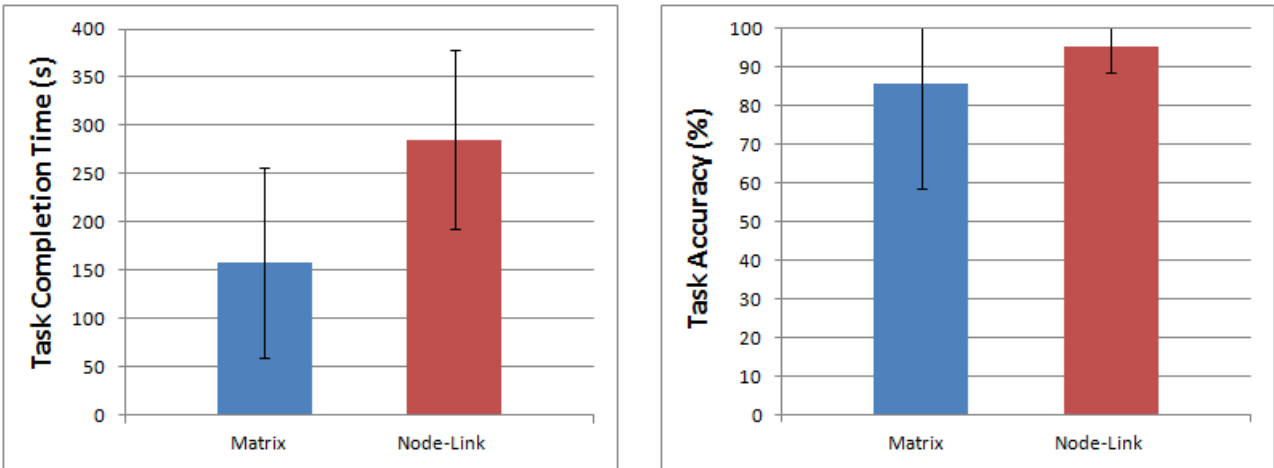


Figure 3. *Left*: Total task completion times (in seconds) on the node-finding task for the matrix visualization and the node-link visualization. Error bars represent +/- one standard deviation from the mean. *Right*: Task accuracy on the node-finding task. Error bars again represent +/- one standard deviation.

3.5 Summary of Findings

Overall, the matrix visualization generally proved the more effective (see summary in Table 1 below). It resulted in 56% faster task completion times on the link-finding task, and 45% faster task completion times on the node-finding task, both of which were statistically significant differences. Differences in accuracy between the two visualization types were non-significant on all four tasks. In terms of consistency of measures, the matrix view resulted in lower standard deviations for the link-finding task completion times (34 versus 92), for the link-finding task accuracy (4 versus 12), and for the link strength judgments (1.34 versus 1.84). The opposite pattern was observed, in which the node-link visualization resulted in more consistent measures, for the node interconnection judgments (9.4 versus 9.6), for the node-finding task completion times (92 versus 99), and for the node-finding accuracy (7 versus 27).

Table 1. A comparative summary of performance measurements. Shaded cells indicate which visualization produced better performance with a statistically significant difference on a particular task (alpha=.05 level).

Task	Measure	Matrix w/ Heat Map	Node-Link Diagram
Link-finding	Time	133 s	303 s
	Accuracy	96 %	94 %
Link Strength	Accuracy	± 6.74 units	± 6.19 units
Node Connectivity	Accuracy	93 %	87 %
Node-finding	Time	158 s	286 s
	Accuracy	86 %	95 %

4. CONCLUSIONS, AND FUTURE WORK

In general, our results support those of Ghoniem, Fekete, & Castagliola [7] who found that for “large” networks (those consisting of at least 20 nodes; 30 nodes were used in this study) the matrix-based visualization resulted in better performance as measured by task completion time and accuracy of judgments. Their one exception in which the node-link diagram was better than the matrix visualization was on a path-tracing task where the shortest path between two nodes had to be identified visually. Their explanation, which we agree with due to our personal experience in working with this visualization type, was that finding lengthy paths (nodes connected by multiple links) can be very difficult to accomplish with the matrix-based visualization; and sometimes is seemingly impossible. Although our experiment did not specifically replicate this condition, the closest comparable condition in our experiment was the node-finding task which essentially required finding a path of length 2 between two given nodes. Under this condition, we found that the matrix visualization resulted in significantly quicker completion times but with comparable accuracy, and with higher variability on both measures. Perhaps this high variability lends some support to Ghoniem, Fekete, & Castagliola’s previous finding that *some* people may struggle a bit more with this particular type of task (even though performance was faster overall with the matrix visualization).

One aspect of network visualization that was not explicitly studied by Ghoniem, Fekete, & Castagliola [7] was the issue of working with weighted network data. The one task category in our study that specifically tested the readability of weighted network data was the link strength estimation task, in which participants judged the relative strength of links

between two given nodes on a scale of 0 to 100. We did not find a significant difference on this task between the two visualization types, although the node-link diagram was just slightly more accurate but exhibited more variability in judgment accuracy. Further research regarding visualizing weighted network data may help to discern performance differences between matrix-based and node-link visualizations, if any exist.

Another key area of future research, which we did not test but which was also mentioned by Ghoniem, Fekete, & Castagliola [7], was the issue of visualizing asymmetric network data, in which the connection from say A to B is a different weight than the connection from B to A. Understanding of this type of data would potentially benefit from the matrix-based visualization, as symmetric network data is redundantly presented via the matrix on each side of the diagonal. Matrix visualizations could be easily adapted for asymmetric network data, but this type of data can prove very difficult for node-link diagrams to portray (since the number of links may have to potentially double to denote all link asymmetries).

One reason that matrix visualizations may offer a particular advantage, at least in the case of our study is this: To do several of the tasks correctly, the participants must first find at least one task-defined target node. In the matrix visualization, this was easy since the nodes were arranged in numerical order (and even if they were in a jumbled order, they would have been neatly arranged along a relatively small strip of space). But for the node-link diagram, finding a particular node is a relatively difficult visual search task, especially in a cluttered presentation. This can be remedied, at least partially, by arranging nodes in patterns that facilitate search, such as in a circular arrangement (somewhat analogous to Facebook's "Friend Wheel" [8]) that could either be ordered or unordered. Unordered and ordered visualizations of the same network data used in the present study are shown in Figure 4 to demonstrate these ideas.

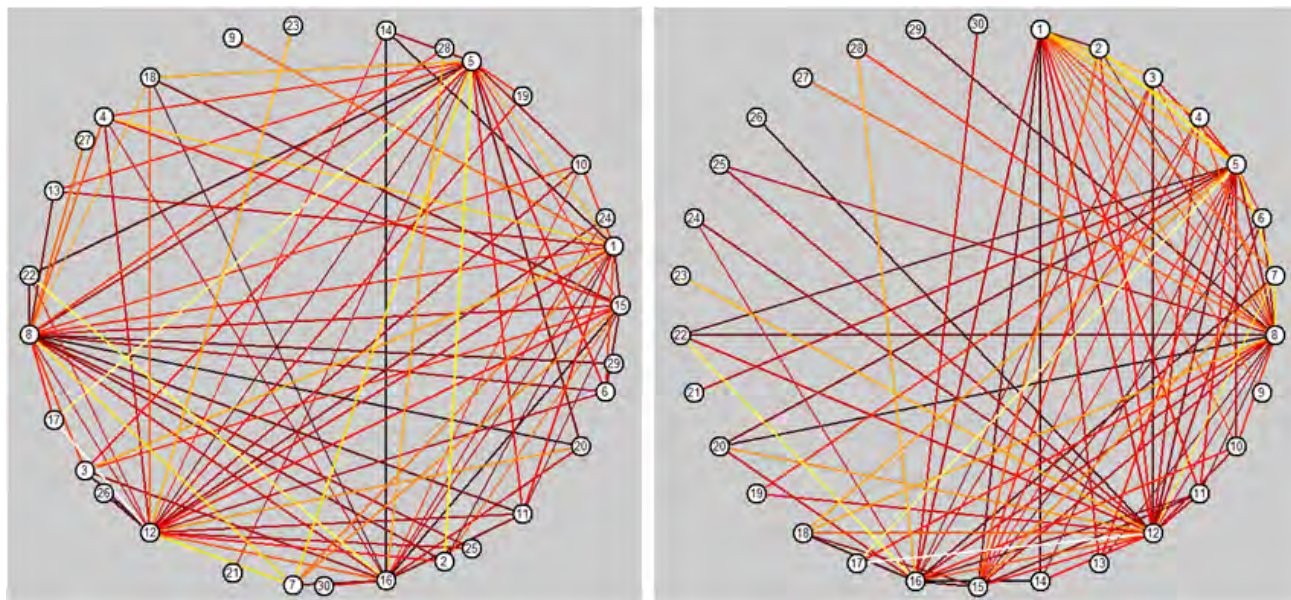


Figure 4. *Left*: An unordered circular arrangement of nodes to facilitate node search and to minimize the effects of link-overlap and clutter. *Right*: The same network with a sequentially ordered circular arrangement.

Ghoniem, Fekete, & Castagliola [7] noted that performance in general (across visualization types) was hampered by increases in size and density of the network graphs; again, this touches on the pervasive problem of scalability of network visualizations (and visualizations in general). Future work on this problem is important so that the visualization community can explore solutions to working with the ever-growing large network datasets pervasive in so many fields, like social network analysis, genetics, biochemistry, intelligence, cybersecurity, and communications. One of the main goals of this research is to determine better, more effective, and more useful ways of presenting information to users for improved data readability and performance in their respective fields.

5. ACKNOWLEDGMENTS

We wish to thank the Tec^Edge Summer at the Edge (SATE) program for its invaluable support, particularly that provided by Dr. Rob Williams from SATE. Thanks also to the participants from the Air Force Research Laboratory's 711th Human Performance Wing for volunteering their time. This work was primarily conducted by the Information Visualization Team at Tec^Edge's Summer at the Edge student program, under the guidance of the Air Force Research Laboratory's 711th Human Performance Wing Battlespace Visualization branch, over 10 weeks in the summer of 2011.

REFERENCES

- [1] Becker, R. A., Eick, S. G., & Wilks, "Visualizing network data," *IEEE Transactions on Visualization and Computer Graphics*, 1(1), 16-21, March (1995).
- [2] Herman, I., Melancon, G., & Marshall, M. S., "Graph visualization and navigation in Information Visualization: A survey," *IEEE Transactions on Visualization and Computer Graphics*, 6(1), 24-43, January-March (2000).
- [3] Purchase, H. C., "The effects of graph layout," Proceedings of the Australasian Computer Human Interaction Conference, p.80, November 29-December 12 (1998).
- [4] Purchase, H. C., Carrington, D., & Alder, J.-A., "Empirical evaluation of aesthetics-based graph layout," *Empirical Software Engineering*, 7, 233-255 (2002).
- [5] Purchase, H. C., Cohen, R. F., & James, M. I., "An experimental study of the basis for graph drawing algorithms," *Journal of Experimental Algorithmics*, 2, 4 (1997).
- [6] Ware, C., Purchase, H., Colpoys, L., & McGill, M., "Cognitive measurements of graph aesthetics," *Information Visualization*, 1(2), 103-110, June (2002).
- [7] Ghoniem, M., Fekete, J.-D., & Castagliola, P., "A comparison of the readability of graphs using node-link and matrix-based visualizations," IEEE Symposium on Information Visualization, 17-24, October 10-12, Austin, Texas (2004).
- [8] Fletcher, T., "Facebook Friend Wheel," Thomas Fletcher [webpage] (2007). Available online at <http://thomas-fletcher.com/friendwheel/>

What is 3D Good For?

A Review of Human Performance on Stereoscopic 3D Displays

John P. McIntire*, Paul R. Havig, & Eric E. Geiselman

711th Human Performance Wing / RHCW, US Air Force Research Laboratory, Dayton, OH, USA

ABSTRACT

This work reviews the human factors-related literature on the task performance implications of stereoscopic 3D displays, in order to point out the specific performance benefits (or lack thereof) one might reasonably expect to observe when utilizing these displays. What exactly is 3D good for? Relative to traditional 2D displays, stereoscopic displays have been shown to enhance performance on a variety of depth-related tasks. These tasks include judging absolute and relative distances, finding and identifying objects (by breaking camouflage and eliciting perceptual “pop-out”), performing spatial manipulations of objects (object positioning, orienting, and tracking), and navigating. More cognitively, stereoscopic displays can improve the spatial understanding of 3D scenes or objects, improve memory/recall of scenes or objects, and improve learning of spatial relationships and environments. However, for tasks that are relatively simple, that do not strictly require depth information for good performance, where other strong cues to depth can be utilized, or for depth tasks that lie outside the effective viewing volume of the display, the purported performance benefits of 3D may be small or altogether absent. Stereoscopic 3D displays come with a host of unique human factors problems including the simulator-sickness-type symptoms of eyestrain, headache, fatigue, disorientation, nausea, and malaise, which appear to effect large numbers of viewers (perhaps as many as 25% to 50% of the general population). Thus, 3D technology should be wielded delicately and applied carefully; and perhaps used only as is necessary to ensure good performance.

Keywords: stereopsis, three-dimensional display, human factors, depth perception, binocular vision

1. INTRODUCTION

Westheimer (1994) observed that most creatures in the animal kingdom, though possessing two or more eyes, lack stereoscopic depth perception and instead have eyes on the sides of their heads pointing outward, giving nearly panoramic vision. But primate visual fields point forward and largely overlap, affording the ability to see stereoscopically. Evolutionarily, there must have been compelling benefits to possessing stereo vision (Fielder & Moseley, 1996); thus, there should be compelling reasons for using 3D displays. But what are they?

Relative to traditional, flat-panel two-dimensional (2D) displays, stereoscopic displays have been shown to enhance performance on a variety of depth-related tasks. These tasks generally include judging absolute and relative distances; finding, identifying, and classifying objects (by breaking camouflage and eliciting perceptual “pop-out”); performing spatial manipulations of objects (object positioning, orienting, and tracking); and navigating. More cognitively, stereoscopic displays can improve the spatial understanding of 3D scenes or objects, improve recall of scenes or objects, and improve learning/training of spatial relationships.

Stereoscopic viewing devices have been around since the time of Wheatstone’s explanation of binocular vision (Wheatstone, 1838) and subsequent invention of the first 3D display, the “stereoscope.” The novelty of the technology and the compelling perceptual experience of depth elicited by stereoscopic displays have fascinated viewers ever since. Despite several disappointing attempts over the years to push the technology into the mainstream, the current resurgence of 3D may suggest it is here to stay. Stereo 3D display technology is now finding wide interest and application in entertainment (especially movies and video-games), medicine (training, imaging, virtual therapy, robotic and teleoperative surgery), industrial design (3D CAD), education and research (scientific and information visualization), and in the military (training, planning, simulation, image analysis, command & control, teleoperative robotics, and unmanned vehicle control).

* contact author: john.mcintire@wpafb.af.mil

In this review, we examine what tasks stereoscopic 3D displays may or may not be well-suited for. We provide some prominent examples to highlight these issues, and we focus mainly on the performance implications--as opposed to the perceptual implications for users. There is a sizable literature, and several reviews, comparing performance on 2D versus 3D displays (e.g., Bemis, Leeds, & Winer, 1988; Haskell & Wickens, 1993; Wickens, 2000; St. John, Cowen, Smallman, & Oonk, 2001; see reviews by Naikar, 1998, and Dixon, Fitzhugh, & Aleva, 2009), but nearly all these works utilize flat 2D displays to show 2D images of 3D perspective geometry (sometimes called 2½D). These are not literally true depth displays in the same sense that stereoscopic displays are 3D because they do not require the binocular visual system. There is also a vast literature on the topic of stereoscopic 3D displays and their implications for *perception* (e.g., Getty, 1982; Patterson & Martin, 1992; Wann, Rushton, & Mon-Williams, 1995; Reinhardt-Rutland, 1996; Pastoor & Wöpking, 1997; Holliman, 2005; IJsselstein, Seuntiëns, & Meesters, 2006; Westheimer, 2011). Additionally, there is experimental psychology literature comparing monocular (one-eyed) versus binocular (two-eyed) vision on real-world exteroceptive and visuomotor task performance (e.g., Jones & Lee, 1981).

Representative summaries on the relationship between *3D stereoscopic displays* and *performance* outside of the medical community appear to be lacking (some reviews of performance with stereoscopic displays in the medical domain are given by Hofmeister, Frank, Cuschieri, & Wade, 2001; Getty & Green, 2007; van Beurden et al., 2009; Held & Hui, 2011). One possible exception is the review of factors affecting human performance in teleoperated robotics by Chen, Haas, & Barnes (2007). They briefly reviewed a half dozen studies comparing stereoscopic displays relative to 2D, and they argue for the use of stereo 3D to improve depth perception, avoid obstacles, and improve teleoperator manipulation. Another brief review by Naikar (1998) argued the opposite, saying that perspective (2½D) displays are usually more appropriate for applied situations because stereoscopic 3D displays are useful only when the display is static or slowly changing, or when monocular cues are poor/absent. Boff (1982) also briefly reviewed 2D versus 3D display performance studies over the preceding 30 years, and noted that “these studies did not find evidence to support a hypothesis of improved visual performance for various applications of 3-D displays...[but] careful review of these past studies suggests many methodological difficulties that raise questions about the validity of the results.” Furthermore, Boff (1982) argued that “the relative effectiveness of 3D displays as compared with encoded volumetric information in two-dimensional (2D) presentations *needs to be determined*” [emphasis added]. Thirty years later, we share this author’s concern, and we believe a more comprehensive review of various human factors studies comparing 2D to 3D is warranted at this time, given the explosion of interest in 3D by both the research community and the general public over the last few years.

2. METHODS

A total of 71 experiments that specifically investigated performance using 2D displays versus stereoscopic 3D displays were found in the human factors literature. Only experiments that included an objective measure of performance were included. Some studies had multiple experiments that were considered individually. All experiments in the medical domain were excluded, though our results will be discussed in comparison to previous medical reviews where appropriate.

Each experiment was classified into one of six different types depending upon what type of performance was primarily being studied: (1) judgments of position and/or distances, (2) finding/identifying/classifying objects, (3) real/virtual spatial manipulations of objects, (4) navigation, (5) spatial understanding/memory/recall, and (6) learning/training/planning. It can be argued whether this classification is too artificial since doing tasks like navigating partially requires judgments of distance and relative object positioning, and manipulating objects generally requires good spatial understanding of the scene. However, we will see that these distinctions, while perhaps not perfect, are potentially useful and important to consider. It should be noted that some studies included multiple experiments; these are broken down by task-type where appropriate.

The 71 experiments were also classified into one of three types depending on their major results: (1) 3D was found to be clearly better than 2D on the main performance measurements, (2) mixed results, in that 3D performance was better on some measurements but not others, and so there was no clear overall “winner,” or (3) 3D performance was found to be statistically indistinguishable from 2D performance (or in very rare cases, worse than 2D).

3. RESULTS

Our results are presented in Tables 1 and 2. Overall, we found that 41 (58%) of the experiments showed 3D to be better than 2D regarding the performance measurements of interest; ten (14%) showed mixed results; and 20 (28%) of the experiments showed no benefit to using 3D. Thus, 72% of the studies showed at least *some* benefit to using stereoscopic displays. Stereoscopic 3D displays showed the clearest benefits during spatial manipulation tasks and during spatial understanding tasks. For the 27 spatial manipulation tasks, 85% of these experiments showed at least some beneficial effect of 3D stereoscopic displays on teleoperative performance. For the thirteen spatial understanding tasks, 92% of the experiments show that 3D offers at least some benefit to performance.

It is interesting and somewhat surprising to note that the beneficial effects of 3D displays are not as strongly apparent for the four other remaining tasks (judging positions/distances, navigation, finding/identifying/classifying objects, and learning/training/planning). In twelve experiments involving the judgment of positions and/or distances, five of the experiments found a clear benefit to using 3D while six of the experiments found no difference between 2D and 3D, with one study showing mixed results. So, for judging positions/distances, half of the experiments found some benefit to using 3D, the other half did not. This same pattern was found for the eight experiments involving navigation tasks (four experiments found benefits to 3D, four did not). Few experiments (by our count, only nine) have compared 2D to 3D displays on tasks involving visually finding, identifying, and/or classifying objects. Again, these results showed about half of the studies were positively affected by 3D (four showed 3D to be beneficial, four others showed no benefit, and one was mixed). We found only two studies outside of the medical literature in which the effects of stereoscopic displays were examined relative to 2D on learning, training, or planning tasks, and both with mixed or negative results. It should be noted that in the medical literature, 3D has been relatively well-studied for the purposes of improving learning (e.g., anatomy), training (e.g., surgical procedures), and planning (e.g., pre-operative strategizing). Next, we discuss our findings within each of these six task categories.

3.1 Spatial Manipulations of Real or Virtual Objects.

In terms of task types, 27 of the 71 experiments (38%) involved performance measurements of teleoperative-type tasks where either real or virtual objects were spatially manipulated through control of some manual interface. It was here that 3D showed its true benefit, as 18 of the 27 experiments (67%) showed a clear benefit to using stereoscopic 3D displays, five experiments (19%) showed a mixed benefit, and only four experiments (15%) showed no benefit of 3D. In other words, fully 85% of the 27 experiments showed at least some beneficial effect of 3D stereoscopic displays on teleoperative performance. And again, these results exclude the many (20 or more) studies of medical 3D teleoperation.

Perhaps it is not surprising to have found that 3D greatly benefits spatial manipulations of objects. There is a substantial literature in the fields of experimental psychology and vision science showing the benefits of binocular vision for real-world reaching, grasping, and manual control (prehension) tasks (e.g., Servos, Goodale, & Jakobson, 1992; Watt & Bradshaw, 2003; Bradshaw, Elliott, Watt, Hibbard, Davies, & Simpson, 2004; Loftus, Servos, Goodale, Mendarozqueta, & Mon-Williams, 2004; Melmoth & Grant, 2006; Melmoth, Finlay, Morgan, & Grant, 2009; Jones & Lee, 1981). However, we omitted these studies from this review because they do not specifically test 2D versus 3D *display* of information; instead, they primarily tested monocular versus binocular viewing of real-world stimuli with no involvement of stereoscopic displays. Given this literature, we had good reason to expect *a priori* that 3D displays might improve the spatial manipulation of objects (either virtual objects or real-world objects accessed via a telemanipulator).

Some of the representative experimental results showing a beneficial effect of 3D displays on spatial manipulation tasks will be discussed next. Pepper, Smith, & Cole (1981) tested control of an undersea telemanipulator and found superior performance for the 3D condition. This was especially true when scene complexity and ambiguity was high. A previous study using a similar telemanipulator task by Pepper, Cole, Merritt, & Smith (1978) also found a benefit of 3D. Kim, Ellis, Tyler, Hannaford, & Stark (1987) showed that performance on a three-axis manual tracking task was generally superior when using 3D. Rosenberg (1993) tested a virtual object placement task and discovered a ten-fold increase in accuracy with the addition of stereoscopy. Hubona, Shirah, & Jennings (2004) showed that stereopsis consistently provided better performance on positioning and resizing of virtual objects in computer-generated spatial tasks. Barfield, Hendrix, & Bystrom (1999) tested virtual path-tracing performance and found that the addition of stereopsis beneficially reduced the time-on-task.

Table 1. Primary results of the literature review for comparing 2D versus 3D stereoscopic displays on performance of various tasks (excluding medical-related studies). The task classification scheme is listed across the top row. Some studies are listed multiple times due to the presence of multiple experiments described in their work. See text for further details.

Authors	Year	Judgments of Position and/or Distances	Finding/Identifying/Classifying Objects	Real/Virtual Spatial Manipulations of Objects	Navigation	Spatial Understanding, Memory, Recall	Learning/Training/Planning
Pepper, Cole, Merritt, & Smith (1)	1978	3D is better					
Singer, Ehrlich, Cinq-Mars, & Papin (1)	1995	3D is better					
Barfield & Rosenberg	1995	3D is better					
Merritt, CuQlock-Knopp, et al.	2005	3D is better					
Reising & Mazur (1)	1990	3D is better					
Hendrix & Barfield	1995	mixed					
Hudson & Cupit	1968	no difference					
Ntuen, Goings, Reddin, & Holmes (1)	2008	no difference					
Ntuen, Goings, Reddin, & Holmes (2)	2008	no difference					
Reising & Mazur (2)	1990	no difference					
Reising & Mazur (3)	1990	no difference					
Willemssen, Gooch, et al.	2008	no difference					
Merritt, CuQlock-Knopp, & Myles	1997		3D is better				
Perlow & Steinberg	1995		3D is better				
Steinberg	1992		3D is better				
Watkins, Heath, et al.	2001		3D is better				
McKee, Watamaniuk, et al.	1997		mixed				
Drasic & Grodski (1)	1993		no difference				
Peinsipp-Byma, Rehfeld, & Eck	2009		no difference				
Steiner & Dotson	1990		no difference				
Zeidner, Sadacca, & Schwartz	1961		no difference				
Pepper, Cole, Merritt, & Smith (2)	1978			3D is better			
Drasic & Grodski (2)	1993			3D is better			
Barfield, Hendrix, & Bystrom	1999			3D is better			
Cole & Parker	1989			3D is better			
Draper, Handel, Hood, & Kring (3)	1991			3D is better			
Dumbreck, Abel, & Murphy	1990			3D is better			
Kim, Ellis, et al.	1987			3D is better			
Lee & Katafiaz	1997			3D is better			
Lee, Park, & Park	1996			3D is better			
McWhorter, Hodges, & Rodriguez	1991			3D is better			
Merritt, Cole, & Ikehara	1991			3D is better			
Pepper, Smith, & Cole	1981			3D is better			
Rosenberg	1993			3D is better			
Scribner & Gombash	1998			3D is better			
Smith, Cole, Merritt, & Pepper	1979			3D is better			
Spain	1990			3D is better			
Spain & Holzhausen	1991			3D is better			
van Beurden, Kuijsters, & IJsselstein	2010			3D is better			
Draper, Handel, Hood, & Kring (2)	1991			mixed			
Drasic (1)	1991			mixed			
Hubona, Shirah, & Jennings	2004			mixed			
Kim, Tendick, Stark	1987			mixed			
Richard, Hareux, Coiffet, & Burdea	1998			mixed			
Singer, Ehrlich, Cinq-Mars, & Papin (2)	1995			no difference			
Draper, Handel, Hood, & Kring (1)	1991			no difference			
Kama & DuMars	1964			no difference			
Park & Woldstad	2000			no difference			
Chen, Oden, Drexler, & Merritt	2010				3D is better		
CuQlock-Knopp, et al.	1995				3D is better		
Merritt & CuQlock-Knopp	1991				3D is better		
Parrish & Williams	1990				3D is better		
Singer, Ehrlich, Cinq-Mars, & Papin (3)	1995				no difference		
de Vries & Padmos	1997				no difference		
de Vries & Padmos	1998				no difference		
Reising & Mazur (4)	1990				no difference		
Aitsiselmi & Holliman	2009					3D is better	
Bourgeois, et al.	2005					3D is better	
Hubona, Shirah, & Fout	1997					3D is better	
Neubauer, Bergner, & Schatz (1)	2010					3D is better	
Sollenberger & Milgram	1993					3D is better	
van Beurden, IJsselstein, & de Kort	2011					3D is better	
Ware & Franck	1996					3D is better	
Ware & Mitchell	2005					3D is better	
Wickens, Merwin, & Lin	1994					3D is better	
Yeh & Silverstein	1992					3D is better	
Brown & Gallimore	1995					mixed	
Lee, MacLachlan, & Wallace	1986					mixed	
Gallimore & Brown	1993					no difference	
Drasic (2)	1991						mixed
Neubauer, Bergner, & Schatz (2)	2010						no difference

Table 2. Summary results of the literature review for comparing 2D versus 3D stereoscopic displays on performance of various tasks (excluding medical-related studies). The task classification scheme is listed across the top row. Numbers represent frequency counts of the experiments, unless otherwise noted. See text for further details.

	Judgments of Position and/or Distances	Finding/ Identifying/ Classifying Objects	Real/Virtual Spatial Manipulations of Objects	Navigation	Spatial Understanding, Memory, Recall	Learning/ Training/ Planning	Totals	% of Totals
3D is better	5	4	18	4	10	0	41	58%
mixed	1	1	5	0	2	1	10	14%
no difference	6	4	4	4	1	1	20	28%
Totals	12	9	27	8	13	2	71	

There were a variety of other telemanipulator-type tasks that also showed a clear benefit to using 3D displays (for further analysis see Table 1 and the references section). From these works it is clear that 3D displays can offer a benefit for telemanipulation or virtual-spatial manipulation tasks. However, not all tasks showed a definite and clear benefit for using 3D over 2D. Kim, Tendick, & Stark (1987) found that 3D generally improved telemanipulator performance, but also found that providing clear or enhanced monoscopic depth cues could create comparable performance advantages. On a teleoperative tapping task, Draper, Handel, Hood, & Kring (1991) showed that 3D was slightly better than 2D only under the most difficult task conditions; for the easier conditions, 3D was not apparently helpful. Several other studies have found similar results in which 2D provided little or no benefit over 3D, and for similar reasons (again, see Table 1).

3.2 Spatial Understanding/Recall/Memory.

The other task-type for which 3D stereo displays showed a very clear overall benefit was spatial understanding, memory, and/or recall tasks. Most of the human factors literature in this category was simply studying spatial understanding, such as the readability of complex graphics or networks (e.g., Ware & Franck, 1996; Ware & Mitchell, 2005; Wickens, Merwin, & Lin, 1994). Out of the thirteen experiments in this category, ten showed a clear benefit of 3D (77%), two showed a mixed result (15%), and one showed no difference (8%). So for spatial understanding tasks, 92% of the experiments showed that 3D offers at least some benefit.

Yeh & Silverstein (1992) showed that the addition of stereoscopy improved judgments of three-dimensional spatial relationships of objects and provided an even larger benefit when there were poor and/or ambiguous monocular depth cues. An experiment by Ware & Franck (1996) discovered that an abstract data graph could be enlarged by a factor of 1.6 and still be understandable if a stereoscopic display was utilized (and could be enlarged by a factor of three if stereopsis was paired with head-coupling). Similarly, Ware & Mitchell (2005) and Wickens, Merwin, & Lin (1994) found that stereoscopic 3D displays improved complex 3D graph reading performance. Stereo 3D displays also help for mental rotation tasks by improving response times and decreasing error rates (Hubona, Shirah, & Fout, 1997; Neubauer, Bergner, & Schatz, 2010).

But 3D displays were not necessarily helpful on all spatial understanding tasks. Lee, MacLachlan, & Wallace (1986) found that stereo helped on one data interpretation task (improved accuracy and response times in reading 3D scatterplot data), but 3D did not help on another task (understanding 3D block diagrams of semi-discrete data, in comparison to a familiar tabular presentation of the same data). In this case, it is possible that the familiarity users had with reading the data in table form suppressed any benefit that might have been provided by the 3D presentation of the block data. Brown & Gallimore (1995) tested participants on 3D-CAD object understanding tasks and showed that 3D was sometimes beneficial, but only when other monocular depth cues were degraded or absent. The one study that showed no benefit of 3D for spatial understanding was performed by these same researchers (Gallimore & Brown, 1993), who previously found that in discriminating between two 3D CAD models, disparity was apparently not needed for performing the task at high-performance levels because adequate monocular cues were present.

3.3 Judging Positions/Distances.

Ogle (1950) argued that a primary purpose for binocular vision is the accurate spatial localization of objects in relation to oneself in the world. Indeed, evidence suggests that spatial localization and distance estimation are improved with binocular vision and that the benefit decreases over larger distances (as disparity cues get relatively smaller; Ciuffreda, 2005; Cutting & Vishton, 1995). Thus, we would expect to see 3D stereoscopic displays improve judgments of object positions and/or distances in space, especially within the near-field of the viewer. Somewhat surprisingly, we found that of the twelve studies testing such judgments, only five showed a clear benefit of 3D, while one study showed a mixed benefit, and six others showed no benefit of 3D over 2D.

Barfield & Rosenberg (1995) found that stereoscopy reduced response times and increased accuracy for depth and altitude judgments of virtual objects. Singer, Ehrlich, Cinq-Mars, & Papin (1995) showed that short range distance estimation was improved with a 3D display. Merritt, CuQlock-Knopp, Kregel, Smoot, & Monaco (2005) found that distance perception of terrain drop-offs was improved with 3D using imagery from teleoperations of ground vehicles. Pepper, Cole, Merritt, & Smith (1978) tested depth discrimination judgments on a virtual two-rod perceptual task and also found that 3D improved performance.

But again, 3D did not always provide a clear benefit over 2D for judgments of positions or distances. In a series of experiments using an airspace disambiguation task, Reising & Mazur (1990) found 3D to be beneficial only when other monocular depth cues were absent. For similar reasons, Ntuen, Goings, Reddin, & Holmes (2008) found no benefit of 3D for virtual object depth judgment tasks in two experiments. Willemsen, Gooch, Thompson, & Creem-Regehr (2008) showed that distance judgments in a virtual environment were comparable over the 2D and 3D conditions. Again, it seems in these cases that the presence of other (monocular) depth cues provided for sufficient performance in the 2D conditions, resulting in mixed or absent benefits of stereoscopic 3D.

3.4 Finding, Identifying, and/or Classifying Objects.

Several studies have shown that conjunctive visual searches can be performed faster (in parallel or “efficiently”) when one of the visual features is stereoscopic depth (e.g., Nakayama & Silverman, 1986; Steinman, 1987). O’Toole & Walker (1997) showed that visual search for a target defined solely by stereoscopic depth was efficient, especially if the target appeared in “front” of distractors located on the fixation plane. Using eye metrics (first saccade accuracy) instead of visual search times, McSorley & Findlay (2001) found that search for a target defined by either a distinct feature or by stereoscopic depth was efficient (although the conjunction of feature plus depth was less efficient). A review of binocular versus monocular sensitivities by Blake, Sloane, & Fox (1981) suggested that form recognition can be improved via stereopsis. Nakayama, Shimojo, & Silverman (1989) suggested that stereopsis is especially helpful in object identification and delineation when parts of objects may be occluded, as in clutter or camouflage. Remarkably, Julesz (1964) showed that observers could detect shapes in particular stimuli when viewed stereoscopically that were *completely invisible* when viewed monoscopically (random dot stereograms). So we have reason to believe that 3D should be helpful on tasks involving the finding, identification, and/or classification of objects. But the experimental studies cited above did not directly test 3D versus 2D on these types of tasks, and so are not directly applicable to our review. Our examination of the literature revealed only nine studies that specifically tested 2D versus 3D on such tasks. We found that only four studies clearly favored 3D displays, while one study was mixed and four others were neutral (or negative). These nine studies will be briefly discussed next.

On a target detection task, in which the targets were camouflaged personnel hidden in visually cluttered terrain sites, Watkins, Heath, Phillips, Valetton, & Toet (2001) discovered that 3D reduced false alarm detection rates by a factor of two. Likewise, using a target detection task with radar imagery in two experiments, Steinberg (1992) and Perlow & Steinberg (1995) found that 3D improved detection performance. And Merritt, CuQlock-Knopp, & Myles (1997) showed that 3D resulted in higher detection rates of critical terrain features for viewers of terrain navigation videos.

There are other studies, however, in which 3D showed a mixed benefit, no benefit, or in at least one case, a detriment. A basic psychophysical study of 3D motion perception by McKee, Watamaniuk, Harris, Smallman, & Taylor (1997) suggested that 3D helped in the detection of static targets in clutter (by eliciting perceptual ‘pop-out’ of the target) but helped very little for detecting straight-moving targets among random-motion distractors. The relatively slow temporal response of the stereo perception system was suggested to be the culprit in this case. Peinsipp-Byma, Rehfeld, & Eck

(2009) tested 2D versus 3D for a variety of image analysis tasks, and found that while 3D seemed to improve mean detection, recognition, and classification times, these differences were not statistically significant. Drasic & Grodski (1993) also found that 2D was comparable to 3D on a teleoperative IED detection task (where other 2D cues to shape/form were present). Zeidner, Sadacca, & Schwartz (1961) tested military image analysts in a between-subjects design, and showed no difference between the 2D and the stereoscopic 3D groups on visual search and identification tasks, in terms of the quality of the information provided by analysts or the confidence assigned to each response. In terms of the number of objects reported in the images, the 2D group tended to identify *more* objects than the 3D group, though it is not clear if this trend was statistically significant. Steiner & Dotson (1990) discovered that the 3D condition was actually *worse* than the 2D condition for a visual search-type task involving the display of tactical aviation information. This was true even though participants seemed to prefer the 3D display format. Clearly, 3D displays are not always beneficial for performance on these types of search, detection, or identification tasks.

3.5 Navigating.

As with the spatial manipulations of objects, there already exists a sizable literature examining monocular versus stereoscopic viewing of real-world stimuli, in this case for the purposes of ground navigation and locomotion (e.g., see Hayhoe, Gillam, Chajka, & Vecellio, 2009; Patla, Niechwiej, Racco, & Goodale, 2002). However, since these studies did not generally test 2D versus 3D displays, but instead tested monocular versus binocular viewing, they were omitted from this review. Our review found that for navigation tasks, a 3D display benefit was found in four of the studies, and was absent in the other four studies.

Merritt & CuQlock-Knopp (1991) showed that using a 3D display for virtual off-road driving produced superior performance compared to the use of a 2D display. They argued that 3D allowed drivers to avoid more terrain hazards and perceive the terrain contours more accurately than 2D. Parrish & Williams (1990) used a simulated hover-in-place helicopter control task and found that 3D clearly provided a benefit over 2D. Performance was best when the stereoscopic display was used in conjunction with other informational displays regarding hovering performance. In a ground navigation task using night vision goggles (NVGs), CuQlock-Knopp, Torgerson, Sipes, Bender, & Merritt (1995) showed that binocular NVGs were better than either biocular or monocular NVGs. On a teleoperative robotic driving task, Chen, Oden, Drexler, & Merritt (2010) found that 3D displays improved performance over 2D displays.

But other studies involving navigation show no benefit of 3D displays. On two simulated unmanned aerial vehicle (UAV) flight tasks, de Vries & Padmos (1997, 1998) found that 3D offered no benefit over 2D for steering accuracy or for various metrics of flight performance (e.g., speed error, route-matching). They speculated that the useful information for completing the task (distance of far objects or waypoints) was simply outside the effective viewing volume of the 3D display. Citing similar reasons, Singer, Ehrlich, Cinq-Mars, & Papin (1995) found no beneficial effect of 3D displays for a virtual room/gate ground navigation task. Reising & Mazur (1990), too, found no benefit of 3D on a Pathway in the Sky virtual aircraft navigation task. Thus, we see that 3D can sometimes be beneficial for navigation tasks, and sometimes not, depending on the task requirements and their relationship to the particular display configurations.

3.6 Learning/Training/Planning.

Surprisingly, we found only two studies in the human factors literature, outside of the medical domain, that specifically studied the effects of 3D displays on learning, training, or planning (Drasic, 1991, and Neubauer, Bergner, & Schatz, 2010). In the medical literature, there appear to be at least a half dozen or more studies involving 3D for helping students to learn anatomy, to train for surgical procedures, train on imagery analysis tasks, or for pre-operative planning. One experiment came close to being included in this task category (Merritt & CuQlock-Knopp, 1991) because they tested perceptual learning of hazard detections in off-road terrain vision, but the researchers only recorded subjective data about learning (and incidentally, found positive results). Drasic (1991) found that for a teleoperative bomb disposal task, the benefit of 3D decayed over time as participants learned how to complete the task effectively using monocular cues and gained familiarity in controlling the telemanipulator. But when the task difficulty was increased, 3D still helped, and was learned fastest. Neubauer, Bergner, & Schatz (2010) found no positive effect of 3D on training outcomes. They studied the mental rotation abilities of males and females both across time (to assess training) and dimensionality (2D versus 3D display). They found that 3D improved reaction times, and that performance got better over time (training effect), but there was no interaction between training and 3D (3D helped, but not differentially across time). The dearth of research in this area suggests a potentially fruitful region for future human factors inquiries.

4. DISCUSSION, CONCLUSIONS, & FUTURE WORK

In summary, 3D seems to help most for depth-related tasks performed in the near-field, especially on difficult/complex tasks (e.g., threading a needle, or visually searching for a small needle in a haystack). In 72% of the 71 studies, 3D showed at least some benefit over 2D viewing. In 58% of the studies, 3D showed a clear and definite benefit. Given our review data, 3D seems to help especially for the spatial manipulations of real or virtual objects (85% of manipulation studies showed some benefit for 3D), and for increasing spatial understanding of complex/ambiguous scenes (92% of spatial understanding studies showed some benefit for 3D). Stereoscopic 3D displays also seems beneficial for tasks involving judging distances, discerning relative positions, finding and identifying objects, and navigating (all of which showed a benefit for 3D on about 50% of the studies involving these tasks). But 3D helps little or sometimes not at all for tasks that are simple or well-learned, or for tasks that do not rely heavily on depth information for good performance. Also, tasks in which other depth cues are strong, or tasks in which depth information is far away or otherwise lies outside the effective viewing volume of the display do not seem to benefit from the use of 3D. We believe that these reasons explain why 28% of the 71 studies found no benefit of 3D over 2D on various performance tasks. For many (if not most) conceivable tasks that people perform on displays, stereo 3D simply may not be necessary to enhance performance.

Overall, our results are somewhat similar to the medical domain reviewed by Hofmeister, Frank, Cuschieri, & Wade (2001). They found that about 50% of the studies showed a significant benefit of stereoscopic 3D displays. And similar to our observations, Hofmeister et al. (2001) noted the effect of task difficulty on the benefit of 3D. For complex tele-surgical maneuvers or in comparison to “incompatible viewing arrangements” (e.g., multiple 2D views), 3D stereoscopic displays were especially beneficial. We also found that 3D was especially helpful for difficult, complex, or unfamiliar depth-related tasks, or for tasks where monocular cues were degraded or absent (these conclusions were consistent with Naikar’s [1998] brief review, as well). Only 13 medical studies were reviewed for objective performance measures at the time by Hofmeister et al. (2001), but with a quick perusal of the present medical literature, we found at least 40 experimental studies pitting 2D against 3D. These studies primarily involve tele-surgical/robotic, medical education/training, and imagery analysis tasks. In contrast to the findings in our review, it appears there are a number of studies in the medical domain which examined the benefit of stereoscopic displays for learning (anatomy), training (surgical procedures), and planning (pre-operative). There also appears to be a large number of studies in the medical literature (but not the general human factors literature) on visually finding/identifying/classifying objects (e.g., analyzing tissue scans, x-rays, etc.) when viewing 2D versus 3D imagery. Apparently, quite a bit of experimental work on stereoscopic displays in the medical community has been conducted over the last few decades. Comprehensive reviews of these newer experiments will be important for both the human factors and medical communities in their future research endeavors.

Stereoscopic 3D displays come with a host of unique human factors problems including the simulator-sickness-type symptoms of eyestrain, headache, fatigue, disorientation, nausea, and malaise, which appear to effect large numbers of viewers. A survey conducted by the American Optometric Association reported that at least a quarter of people who watched 3D films, television, or videogames experienced such symptoms (AOA.org, 2010). And an informal online survey by HomeTheater.com found that 53% of people who have viewed 3D content have experienced these sickness symptoms (Wilkinson, 2011). Thus, perhaps as many as 25% or even 50% of the general population may have uncomfortable experiences when viewing 3D displays. And since stereoscopic 3D displays seem to offer very select benefits on specific (depth-related) tasks, our review suggests that the technology should be wielded delicately and applied carefully, and perhaps used only as is necessary to ensure good performance.

REFERENCES

- [1] Aitsiselmi, Y. & Holliman, N. S. (2009). “Using mental rotation to evaluate the benefits of stereoscopic displays.” *Proceedings of SPIE-IS&T Electronic Imaging, Stereoscopic Displays and Applications XX*, 7237, 72370Q-1.
- [2] American Optometric Association (2010). “3-D TV and movies look to attract viewers but not everyone can ‘see’ what all the hype is about.” [digital press release] on May 18, 2010. Available: <http://www.aoa.org/x15123.xml>
- [3] Barfield, W. Hendrix, C., & Bystrom, K. E. (1999). “Effects of stereopsis and head tracking on performance using desktop virtual environment displays.” *Presence: Teleoperators and Virtual Environments*, 8(2), pp. 237-240.

- [4] Barfield, W., & Rosenberg, C. (1995). "Judgments of azimuth and elevation as a function of monoscopic and binocular depth cues using a perspective display." *Human Factors*, 37, pp. 173-181.
- [5] Bemis, S. V., Leeds, J. L., & Winer, E. A. (1988). "Operator performance as a function of types of display: Conventional versus perspective." *Human Factors*, 30, pp. 163-169.
- [6] Blake, R., Sloane, M., & Fox, R. (1981). "Further developments in binocular summation." *Perception & Psychophysics*, 30, pp. 266-276.
- [7] Boff, K. R. (1982). "Critical research issues on cockpit applications of 3-D displays." In Getty, D. J., *Three-dimensional displays: Perceptual research and applications to military systems*. Proceedings, National Academy of Sciences, National Research Council: Committee on Human Factors, ONR technical report, A122577, 1982.
- [8] Bourgois, M., Cooper, M., Duong, V., Hjalmarsson, J., Lange, M., & Ynnerman, A. (2005). "Interactive and immersive 3D visualization for ATC." Proceedings of the 6th USA-Europe ATM R&D Seminar, Baltimore, MD, USA, June 2005, pp. 303-309 (FAA).
- [9] Bradshaw, M. F., Elliott, K. M., Watt, S. J., Hibbard, P. B., Davies, I. R., & Simpson, P. J. (2004). "Binocular cues and the control of prehension." *Spatial Vision*, 17(1-2), pp 95-110.
- [10] Brown, M. E., & Gallimore, J. J. (1995). "Visualization of three-dimensional structure during computer-aided design." *International Journal of Human-Computer Interaction*, 7(1), pp. 37-56.
- [11] Chen, J. Y. C., Haas, E. C., & Barnes, M. J. (2007). "Human performance issues and user interface design for teleoperated robots." *IEEE Transactions on Systems, Man, & Cybernetics – Part C: Applications and Reviews*, 37(6), pp. 1231-1245.
- [12] Chen, J. Y. C., Oden, R. V. N., Drexler, J. M., & Merritt, J. O. (2010). "Evaluation of stereoscopic display for indirect-vision driving and robot teleoperation." *The WSTIAC Quarterly*, 10(2), pp. 1-9.
- [13] Ciuffreda, K. J. (2005). "Why two eyes?" Specialized Visual Solutions [webpage]. Available: <http://www.svs-greensboro.com/why2eyes.php>
- [14] Cole, R. E., Merritt, J. O., Fore, S., & Lester, P. (1990). "Remote-manipulator tasks impossible without stereo TV." Proceedings of SPIE, 1258, pp. 255-265.
- [15] CuQlock-Knopp, V. G., Torgerson, W., Sipes, D. E., Bender, E., & Merritt, J. O. (1995). "A comparison of monocular, biocular, and binocular night vision goggles for traversing off-road terrain on foot." U.S. Army Research Laboratory, Technical Report ARL-TR-747.
- [16] Cutting, J. E., & Vishton, P. M. (1995). "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth." In PERCEPTION OF SPACE AND MOTION, W. Epstein & S. Rogers (eds.), New York: Academic Press.
- [17] de Vries, S. C., Padmos, P. (1997). "Steering a simulated unmanned aerial vehicle using a head-slaved camera and HMD." Proceedings of SPIE, 3058, pp. 24-33.
- [18] de Vries, S. C., Padmos, P. (1998). "Steering a simulated unmanned aerial vehicle using a head-slaved camera and HMD: Effects of HMD quality, visible vehicle references and extended stereos cueing." Proceedings of SPIE, 3362, pp. 80-91.
- [19] Dixon, S., Fitzhugh, E., & Aleva, D. (2009). "Human factors guidelines for applications of 3D perspectives: A literature review." Proceedings of SPIE, 7327, 73270K-1.
- [20] Draper, J. V., Handel, S., Hood, C. C., & Kring, C.T. (1991). "Three experiments with stereoscopic television: When it works and why." Proceedings of the IEEE International Conference on Decision Aiding for Complex Systems – Systems, Man, & Cybernetics, 2, pp. 1047-1052.
- [21] Drasic, D. (1991). "Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing." Proceedings of the Human Factors Society 35th Annual Meeting, pp. 1367-1371.
- [22] Drasic, D., & Grodski, J. J. (1993). "Using stereoscopic video for defence teleoperation." Proceedings of SPIE, 1915, pp. 58-69.

- [23] Dumbreck, A. A., Abel, E., & Murphy, S. (1990). "3-D TV system for remote handling: Development and evaluation." *Proceedings of SPIE, Stereoscopic Displays and Applications*, 1256, pp. 226-236.
- [24] Fielder & Moseley (1996). "Does stereopsis matter in humans?" *Eye*, 10, pp. 233-238.
- [25] Gallimore, J. J., & Brown, M. E. (1993). "Visualization of 3-D computer-aided design objects." *International Journal of Human-Computer Interaction*, 5(4), pp. 361-382.
- [26] Getty, D. J. (1982). "Three-dimensional displays: Perceptual research and applications to military systems." *Proceedings, National Academy of Sciences, National Research Council: Committee on Human Factors*, ONR technical report, A122577.
- [27] Getty, D. J., & Green, P. J. (2007). "Clinical applications for stereoscopic 3-D displays." *Journal of the Society for Information Display*, 15(6), pp. 377-384.
- [28] Haskell, I. D., & Wickens, C. D. (1993). "Two- and three-dimensional displays for aviation: A theoretical and empirical comparison." *International Journal of Aviation Psychology*, 3(1), pp. 87-109.
- [29] Hayhoe, M., Gillam, B., Chajka, K., & Vecellio, E. (2009). "The role of binocular vision in walking." *Visual Neuroscience*, 26(1), pp. 73-80.
- [30] Held, R. T., & Hui, T. T. (2011). "A guide to stereoscopic 3D displays in medicine." *Academic Radiology*, 18(8), pp. 1035-1048.
- [31] Hendrix, C. & Barfield, W. (1995). "Relationship between monocular and binocular depth cues for judgements of spatial information and spatial instrument design." *Displays*, 16(3), pp. 103-113.
- [32] Hofmeister, J. Frank, T. G., Cuschieri, A. & Wade, N. J. (2001). "Perceptual aspects of two-dimensional and stereoscopic display techniques in endoscopic surgery: Review and current problems." *Seminars in Laparoscopic Surgery*, 8(1), pp. 12-24.
- [33] Holliman, N. (2005). "3D display systems." In *HANDBOOK OF OPTOELECTRONICS*. IOP Press, London. Available online at: <http://www.dur.ac.uk/n.s.holliman/Presentations/3dv3-0.pdf>
- [34] Hubona, G. S., Shirah, G. W., & Fout, D. G. (1997). "3D object recognition with motion." *Proceedings of the CHI 97 (Computer-Human Interaction) Electronic Publication (Late-breaking/short talks)*, extended abstract.
- [35] Hubona, G. S., Shirah, G. W., & Jennings, D. K. (2004). "The effects of cast shadows and stereopsis on performing computer-generated spatial tasks." *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 34(4), pp. 483-493.
- [36] Hudson, E., & Cupit, G. (1968). "Stereo TV enhancement study." *NASA Technical Report NAS 8-21201*. Syosset, NY: Kollsman Instrument Corporation.
- [37] IJsselsteijn, W. A., Seuntjens, P. J. H., & Meesters, L. M. J. (2006). "Human factors of 3D Displays." Chapter 12 in Schreer, Kauff, & Sikora (Eds.) *3D VIDEOCOMMUNICATION: ALGORITHMS, CONCEPTS, AND REAL-TIME SYSTEMS IN HUMAN CENTERED COMMUNICATION*. Wiley & Sons, Ltd. DOI: 10.1002/0470022736.ch12
- [38] Jones, R. K., Lee, D. N. (1981). "Why two eyes are better than one: The two views of binocular vision." *Journal of Experimental Psychology: Human Perception and Performance*, 7(1), pp. 30-40.
- [39] Julesz, B. (1964). "Binocular depth perception without familiarity cues." *Science*, 145, pp. 356-362.
- [40] Kama, W. N., & DuMars, R. C. (1964). "Remote viewing: A comparison of direct viewing, 2D and 3D television." *Aerospace Medical Research Laboratory, Technical Document Report, AMRL-TDR-64-15*.
- [41] Kim, W. S., Ellis, S. R., Tyler, M. E., Hannaford, B., & Stark, L. W. (1987). "Quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking tasks." *IEEE Transactions on Systems, Man, and Cybernetics*, 17(1), pp. 61-72.

- [42] Kim, W. S., Tendick, F., & Stark, L. W. (1987). "Visual enhancements in pick-and-place tasks: Human operators controlling a simulated cylindrical manipulator." *IEEE Journal of Robotics and Automation*, RA-3(5), pp. 418-425.
- [43] Lee, B. C., & Katafiaz, M. E. (1997). "Evaluation of a 3D autostereoscopic display for telerobotic operations." *Proceedings of SPIE*, 3012, pp. 48-58.
- [44] Lee, J. M., MacLachlan, J., & Wallace, W. A. (1986). "The effects of 3D imagery on managerial data interpretation." *MIS Quarterly*, 10(3), pp. 257-269.
- [45] Lee, N. S., Park, J. H., & Park, K. S. (1996). "Reality and human performance in a virtual world." *International Journal of Industrial Ergonomics*, 18, pp. 187-191.
- [46] Loftus, A., Servos, P., Goodale, M. A., Mendarozqueta, N., & Mon-Williams, M. (2004). "When two eyes are better than one in prehension: monocular viewing and end-point variance." *Experimental Brain Research*, 158(3), pp. 317-327.
- [47] McKee, S. P., Watamaniuk, S. N. J., Harris, J. M., Smallman, H. S., & Taylor D. G. (1997). "Is stereopsis effective in breaking camouflage for moving targets?" *Vision Research*, 37(15), pp. 2047-2055.
- [48] McSorley, E., & Findlay, J. M. (2001). "Visual search in depth." *Vision Research*, 41, pp. 3487-3496.
- [49] McWhorter, S., Hodges, L., & Rodriguez, W. (1991). "Evaluation of display parameters affecting user performance of an interactive task in a virtual environment." Georgia Institute of Technology. Graphics, Visualization, & Usability Center, GIT-GVU-91-31, November.
- [50] Melmoth, D. R., & Grant, S. (2006). "Advantages of binocular vision for the control of reaching and grasping." *Experimental Brain Research*, 171(3), pp. 371-388.
- [51] Melmoth, D. R., Finlay, A. L., Morgan, M. J., & Grant, S. (2009). "Grasping deficits and adaptations in adults with stereo vision losses." *Investigative Ophthalmology & Visual Science*, 50(8), pp. 3711-3720.
- [52] Merritt, J. O., Cole, R. E., & Ikehara, C. (1991). "A rapid-sequential-positioning task for evaluating motion parallax and stereoscopic 3D cues in teleoperator displays." *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics: Decision Aiding for Complex Systems*, pp. 1041-1046, October.
- [53] Merritt, J. O., & CuQlock-Knopp, V. G. (1991). "Perceptual training with cues for hazard detection in off-road driving." *Proceedings of SPIE, Stereoscopic Displays and Applications II*, 1457, pp. 133-138.
- [54] Merritt, J. O., CuQlock-Knopp, V. G., Kregel, M., Smoot, J., & Monaco, W. (2005). "Perception of terrain drop-offs as a function of L-R viewpoint separation in stereoscopic video." *Proceedings of SPIE*, 5800, pp. 169-176.
- [55] Merritt, J. O., CuQlock-Knopp, V. G., & Myles, K. (1997). "Enhanced perception of terrain hazards in off-road path choice: stereoscopic 3D versus 2D displays." *Proceedings of SPIE*, 3062, pp. 101-104.
- [56] Naikar, N. (1998). "Perspective displays: A review of human factors issues." Aeronautical and Maritime Research Laboratory, Defense Science & Technology Organisation, Australia; technical report DSTO-TR-0630.
- [57] Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). "Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects." *Perception*, 17, pp. 55-68.
- [58] Nakayama, K., & Silverman, G. H. (1986). "Serial and parallel processing of visual feature conjunctions." *Nature*, 320, pp. 264-265.
- [59] Neubauer, A. C., Bergner, S., & Schatz, M. (2010). "Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation." *Intelligence*, 38(5), pp. 529-539.
- [60] Ntuen, C. A., Goings, M., Reddin, M., & Holmes, K. (2008). "Comparison between 2-D & 3-D using an autostereoscopic display: The effects of viewing field and illumination on performance and visual fatigue." *International Journal of Industrial Ergonomics*, 39(2), pp. 388-395.
- [61] Ogle, K. N. (1950). *Researches in Binocular Vision*. New York: Hafner.

- [62] O'Toole, A. J., & Walker, C. L. (1997). "On the preattentive accessibility of stereoscopic disparity: Evidence from visual search." *Perception & Psychophysics*, 59(2), pp. 202-218.
- [63] Park, S. H., & Woldstad, J. C. (2000). "Multiple two-dimensional displays as an alternative to three-dimensional displays in telerobotic tasks." *Human Factors*, 42(4), pp. 592-603.
- [64] Parrish, R. V., & Williams, S. P. (1990). "Stereopsis cueing effects on hover-in-turbulence performance in a simulated rotorcraft." NASA Technical Paper 2980, AVSCOM Technical Report 90-B-002.
- [65] Patla, A. E., Niechwiej, E., Racco, V., & Goodale, M. A. (2002). "Understanding the contribution of binocular vision to the control of adaptive locomotion." *Experimental Brain Research*, 142(4), pp. 551-561.
- [66] Patterson, R. & Martin, W. L. (1992). "Human stereopsis." *Human Factors*, 34(6), pp. 669-692.
- [67] Pastoor, S., & Wöpking, M. (1997). "3-D displays: A review of current technologies." *Displays*, 17, pp. 100-110.
- [68] Peinsipp-Byma, E., Rehfeld, N., & Eck, R. (2009). "Evaluation of stereoscopic 3D displays for image analysis tasks." Proceedings of SPIE, 7237, 72370L.
- [69] Pepper, R. Cole, R., Merritt, J., & Smith, D. (1978). "Operator performance using conventional or stereo displays." *Optical Engineering*, 17(4), pp. 411-415.
- [70] Pepper, R., Smith, D., & Cole, R. (1981). "Stereo TV improves operator performance under degraded visibility conditions." *Optical Engineering*, 20, pp. 579-585.
- [71] Perlow, R. B., & Steinberg, B. D. (1995). "Enhanced target detection using stereoscopic imaging radar." IEEE Transactions on Aerospace and Electronic Systems, 31(3), pp. 1139-1148.
- [72] Reinhardt-Rutland, A. H. (1996). "Remote operation: A selective review of research into visual depth perception." *Journal of General Psychology*, 123(3), pp. 237-248.
- [73] Reising, J. M., & Mazur, K. M. (1990). "3-D displays for cockpits: Where they payoff." Proceedings of SPIE, 1256, pp. 35-43.
- [74] Richard, P., Hareux, P., Coiffet, P., & Burdea, G. (1998). "Effect of stereoscopic viewing on human tracking performance in dynamic virtual environments." Virtual Worlds 98, LNAI 1434, pp. 97-106.
- [75] Rosenberg, L. B. (1993). "The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks." Proceedings of IEEE VRAIS '93, pp. 27-31.
- [76] Scribner, D. R., & Gombash, J. W. (1998). "The effect of stereoscopic and wide field of view conditions on teleoperator performance." Army Research Laboratory, technical report, ARL-TR-1598.
- [77] Servos, P., Goodale, M. A., & Jakobson, L. S. (1992). "The role of binocular vision in prehension: A kinematic analysis." *Vision Research*, 32(8), pp. 1513-1521.
- [78] Singer, M. J., Ehrlich, J., Cinq-Mars, S., & Papin, J.-P. (1995). "Task performance in virtual environments: Stereoscopic versus monoscopic displays and head-coupling." U.S. Army Research Institute, Technical Report 1034.
- [79] Smith, D. C., Cole, R. E., Merritt, J. O., & Pepper, R. L. (1979). "Remote operator performance comparing mono and stereo TV displays: The effects of visibility, learning, and task factors." Naval Ocean Systems Center, technical report, NOSC-TR-380, February.
- [80] Sollenberger, R. L., & Milgram, P. (1993). "Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task." *Human Factors*, 35(3), pp. 483-499.
- [81] Spain, E. H. (1990). "Stereo advantage for a peg-in-hole task using a force-feedback manipulator." Proceedings of SPIE, 1256, pp. 244-254.
- [82] Spain, E. H., & Holzhausen, K.-P. (1991). "Stereoscopic versus orthogonal view displays for performance of a remote manipulation task." Proceedings of SPIE, 1457, pp. 103-110.

- [83] St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). "The use of 2D and 3D displays for shape-understanding versus relative-position tasks." *Human Factors*, 43(1), pp. 79-98.
- [84] Steinberg, B. D. (1992). "Target detection-sensitivity enhancement using high resolution radar and 2-D and 3-D stereo target displays." *IEEE Transactions on Aerospace and Electronic Systems*, 28, pp. 886-890.
- [85] Steiner, B. A., & Dotson, D. A. (1990). "The use of 3-D stereo display of tactical information." Proceedings of the Human Factors and Ergonomics Society Annual Meeting, October, 34(1), pp. 36-40.
- [86] Steinman, S. B. (1987). "Serial and parallel search in pattern vision?" *Perception*, 16, pp. 389-398.
- [87] van Beurden, M. H. P. H., IJsselsteijn, W. A., & de Kort, Y. A. W. (2011). "Evaluating stereoscopic displays: Both efficiency measures and perceived workload sensitive to manipulations in binocular disparity." Proceedings of SPIE-IS&T Electronic Imaging, Stereoscopic Displays and Applications XXII, 7863, 786316-1.
- [88] van Beurden, M. H. P. H., Kuijsters, A., & IJsselsteijn, W. A. (2010). "Performance of a path tracing task using stereoscopic and motion based depth cues." Second International Workshop on Quality of Multimedia Experience, QoMEX 2010, Trondheim, Norway, June 2010.
- [89] van Beurden, M. H. P. H., van Hoey, G., Hatzakis, H., & IJsselsteijn, W. A. (2009). "Stereoscopic displays in medical domains: A review of perception and performance effects." Proceedings of SPIE, 7240, 72400A-1.
- [90] Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). "Natural problems for stereoscopic depth perception in virtual environments." *Vision Research*, 35(19), pp. 2731-2736.
- [91] Ware, C., & Franck, G. (1996). "Evaluating stereo and motion cues for visualizing information nets in three dimensions." *ACM Transactions on Graphics*, 15(2), pp. 121-140.
- [92] Ware, C., & Mitchell, P. (2005). "Reevaluating stereo and motion cues for visualizing graphs in three dimensions." Proceedings of the 2nd Symposium on Applied Perception in Graphics, APGV '05.
- [93] Watkins, W. R., Heath, G. D., Phillips, M. D., Valeton, M., & Toet, A. (2001). "Search and target acquisition: single line of sight versus wide baseline stereo." *Optical Engineering*, 40(9), pp. 1914-1927.
- [94] Watt, S. J., & Bradshaw, M. F. (2003). "The visual control of reaching and grasping: Binocular disparity and motion parallax." *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), pp. 404-415.
- [95] Westheimer, G. (1994). "The Ferrier Lecture, 1992. Seeing depth with two eyes: Stereopsis." Proceedings: Biological Sciences, 257, pp. 205-214.
- [96] Westheimer, G. (2011). "Three-dimensional displays and stereo vision." Proceedings of the Royal Society B, 278, pp. 2241-2248.
- [97] Wheatstone, C. (1838). "On some remarkable, and hitherto unobserved, phenomena of binocular vision." *Philosophical Transactions of the Royal Society of London*, 128, pp. 371-394.
- [98] Wickens, C. (2000). "The when and how of using 2D and 3D displays for operational tasks." Proceedings of the IEA 2000/HFES 2000 Congress, pp. 3-403 – 3-406.
- [99] Wickens, C. D., Merwin, D. H., & Lin, E. L. (1994). "Implications of graphics enhancements for the visualization of scientific data: dimensional integrality, stereopsis, motion, and mesh." *Human Factors*, 36(1), pp. 44-61.
- [100] Wilkinson, S. (2011). "3D Sickness Explained?" Home Theater [web article], 25 January 2011. Available: <http://www.hometheater.com/content/3d-sickness-explained>
- [101] Willemsen, P., Gooch, A. A., Thompson, W. B., & Creem-Regehr, S. H. (2008). "Effects of stereo viewing conditions on distance perception in virtual environments." *Presence*, 17(1), pp. 91-101.
- [102] Yeh, Y., Silverstein, L. D. (1992). "Spatial judgments with monoscopic and stereoscopic presentation of perspective displays." *Human Factors*, 34(5), pp. 583-600.
- [103] Zeidner, J., Sadacca, R., & Schwartz, A. I. (1961). "Human factors studies in image interpretation: The value of stereoscopic viewing." Adjutant General's Office (Army) Washington, D.C., technical report, June 1961.

VAST 2011 Challenge: Cyber Security and Epidemic

Georges Grinstein
University of
Massachusetts
Lowell

Kristin Cook
Pacific Northwest
National Laboratory

Paul Havig
Air Force Research
Laboratory

Kristen Liggett
Air Force Research
Laboratory

Bohdan Nebesh
Department of
Defense

Mark Whiting
Pacific Northwest
National Laboratory

Kirsten Whitley
Department of
Defense

Shawn Konenci
University of
Massachusetts
Lowell

ABSTRACT

The 6th Visual Analytics Science and Technology (VAST) Challenge posed three related mini-challenges for participants to solve using a combination of visual analytics software and their own analytic reasoning abilities. Teams could solve one, two or all three mini-challenges and assess the overall situation to enter the Grand Challenge. Mini-challenge One (MC1) involved the characterization of the spread of an epidemic using given maps, geospatial and text data gathered from microblog tweets. Mini-challenge Two (MC2) involved the development and use of situation awareness data to identify issues of concern in the computer networking operations at a major freight shipping company. Mini-challenge Three (MC3) involved the exploration of a corpus of news articles to examine terrorist threats to a metropolitan area. The Grand Challenge was to determine whether the epidemic spread, the network events, and the potential terrorist groups identified in the mini-challenges were related. Participants were asked to analyze the data and provide solutions and explanations for the various challenges. The Challenge data sets were downloaded by nearly 600 people by the time submissions closed. The Challenge received 56 submissions, drew participants from 11 different countries, and gave 12 varied awards.

Keywords: visual analytics, human information interaction, sense making, evaluation, metrics, contest.

Index Terms: H.5.2 [Information Interfaces & Presentations]: User Interfaces – Evaluation/methodology

1 BACKGROUND

Now in its sixth year, the objective of the VAST Challenge [1] is to provide researchers with realistic tasks and data sets for evaluating their software, as well as to advance the field in solving more complex problems. The VAST Challenge is designed to help researchers understand how their software would be used in a novel analytic task and determine if their data transformations, visualizations, and interactions would be beneficial for particular analytic tasks. Researchers and software providers have repeatedly used the data sets from throughout the life of the VAST Challenge as benchmarks to demonstrate and test the capabilities of their systems. The ground truth that is

embedded in the data sets has helped researchers evaluate and strengthen the utility of their visualizations.

2 VAST 2011 CHALLENGE SCOPE

The VAST 2011 Challenge consisted of three related mini-challenges (MC1, MC2, and MC3) and one Grand Challenge (GC). Each mini-challenge consisted of a data set, instructions, and questions to be answered. The GC required participants to integrate the information from all three data sets and write a brief summary and explanation of the overall situation.

The VAST 2011 scenario featured various identifiable terrorist activities, an epidemic, and a freight company's network security logs. All of the events in the scenario occurred in the fictional city of Vastopolis during the first half of 2011. MC1 consisted of text (tweets) which participants needed to process to identify the symptoms and details of an epidemic. There were two different sets of illnesses, a waterborne illness and an airborne illness. The participants were asked to locate and pinpoint the source of the epidemic, to describe the method of transmission of the epidemic, and determine if deployment of treatment resources outside of the affected area was necessary. MC2 provided over 8GB of network logs, including vulnerability scans, firewall logs, operating system security logs, intrusion detection system logs, and optional packet capture data. Participants were asked to develop a situation awareness visualization encompassing this data and to identify major network events transpiring over a three-day window. MC3 required participants to analyze a corpus of over 4,000 news articles to determine if there were any imminent terrorist threats.

The data for MC1 and MC3 were developed by the IVPR² at the University of Massachusetts Lowell. The first task for the MC1 scenario was the identification of a city with a river running through its center. Next, the tweet data set was created using a mixture of real tweets collected from Twitter along with a set of synthetic tweets generated with controlled tokens and synonyms using simple dictionaries. These were processed to remove foul language and embed map (latitude-longitude) and time information matching associated weather data. MC3's data set was created from a corpus of old news articles filtered to remove proper nouns and other text (dates and unique headers) that would give away the data set's true origin. There were about fifty additional articles injected into the data set that contained both ground truth and secondary misleading scenarios.

The MC2 data set was developed by Pacific Northwest National Laboratory. This data was created by developing a synthetic network which simulated the architecture of the fictitious freight company. The data were produced by simulating activity, including attacks, on this network over the course of the three day operating period.

grinstein@cs.uml.edu; kris.cook@pnnl.gov;
Paul.havig@wpafb.af.mil; Kristen.liggett@wpafb.af.mil;
danko1@gmail.com; mark.a.whiting@pnnl.gov;
skoneci@yahoo.com

² Institute for Visualization and Perception Research

3 VAST 2011 CHALLENGE SUBMISSIONS

Teams were asked to provide a video and a concise process description as to how they arrived at their conclusions and how the various visualizations and tools helped in the analysis.

Participants submitted 5 GC entries and 51 MC entries. Table 1 shows a comparison of the number of submissions over the life of the VAST Challenge.

	2006	2007	2008	2009	2010	2011
MC1 Submissions	-	-	22	22	14	30
MC2 Submissions	-	-	13	17	22	8
MC3 Submissions	-	-	12	5	17	13
MC4 Submissions	-	-	20	-	-	-
GC Submissions	6	7	6	5	5	5
Total Submissions	6	7	73	49	58	56

Table 1. Summary of Number of Submissions by Year

The number of entries this year was on par with those of the past years. This was especially rewarding given the diversity of data included in the challenge. Most interesting is that, as of publication time, the dataset has been downloaded 671 times, as compared to 537 in 2010. In addition, this year's challenge had more than twice the number of student entries as non-student entries (38 vs. 18). To successfully compete in the GC, participants were required to transform, visualize, and analyze data from all three mini-challenges. The analytic tasks were diverse, ranging from situation awareness to identification of geospatial and temporal trends to criminal investigation.

4 REVIEW PROCESS

The VAST Challenge Review Committee recruited reviewers from throughout the visualization and analysis communities. Several subject matter experts learned about the Challenge through a blog entry published by an analyst educator. In all, 56 reviewers participated, with reviewers providing between one and six reviews each. Three to six external peer reviewers, including at least one subject matter expert, reviewed each entry. The reviewers were given an opportunity to recommend submissions for specific awards.

Each reviewer was given electronic access to the solutions for their assigned submissions. Reviewers were asked to rate the analytic process, the visualizations, the interactions, and the novelty of the submission. Reviewers were also asked to evaluate the accuracy of each team's solution. However, as the tasks and data sets for this challenge were more realistically complex, accuracy was not the only measure of interest. For example, the groups and events associated with the terrorist threat of MC3 relied on finding thirteen critical articles in a corpus of over 4,000 articles. To appropriately identify all of the network events embedded in MC2, participants needed to jointly analyze all sources provided and discriminate between innocuous anomalies and important network attacks. In both cases, all teams were able to discover non-trivial information in the data sets and several teams achieved close to accurate solutions. Interestingly, as in the past, some teams found other patterns not anticipated by the developers of the data set.

The VAST Challenge Review Committee held a two-day meeting to determine awards. The Challenge committee members each took responsibility for reading and summarizing the submitted reviews for one or more of the mini-challenges. The committee reviewed and evaluated the award recommendations from the reviewers and identified additional appropriate awards.

As in previous years, the awards were not pre-established. Instead, the committee identified awards recognizing the best qualities in the submissions. Awards were given for overall quality, analytic processes, innovative approaches, clarity of explanation, and potential for scalability. All teams receiving an award were given the opportunity to contribute two-page summaries for the proceedings. As in the past, all submissions and publications will be available at the Visual Analytics Benchmark Repository [2]. All teams will receive certificates of participation and are invited to the VAST Challenge Participants' Workshop at the 2011 IEEE VisWeek Conference to demonstrate their software and approach.

5 SUMMARY OF VAST CHALLENGE 2011 AWARDS

Several trends were noted in this year's submissions. While a number of teams wrote custom software to address the challenges, as had occurred in the past challenges, several teams developed software using visualization toolkits. It was common for teams to use existing software, including commercial software, to address all or parts of the Challenge. Of particular interest was the fact that a tool that was used for a previous year's Challenge as a research prototype was used in this year's challenge by several teams as an established "off-the-shelf" tool.

Also notable in this year's Challenge were a few entries that took advantage of new form factors for innovative analysis. One entry made use of a tablet device, while at least two others made use of very large display environments.

The level of data preprocessing performed by the teams was notable on MC2. Although the data provided for this mini-challenge was relatively small compared to real-world environments, the size and diversity of the data sources necessitated that teams develop strategies for data management and multi-type data analysis. In the previous Challenges potential submitters asked the VAST Challenge committee to provide preprocessing in the form of extracted entities, text processing similar to that required for MC1 and MC3. This year no such requests were received.

MC2 asked submitters to provide situation awareness visualizations, which would permit users to see at a glance the health of their network and the presence of emerging issues. However, most of the submissions provided visualizations and interactions more oriented toward forensic analysis than situation awareness. This represents an opportunity for further development and future Challenge tasks.

Awards were given for the novel use of specific tools (for example, the use of word clouds for filtering other visualization), for outstanding analysis, for novel extensions to mobile devices and to large screen workspaces to support collaboration, for informative use of statistics and evidence in a report, for innovative tool adaptation, and for scalability.

Challenge	Student Team	Non-Student Team	Total Awards
MC1	3	2	5
MC2	3	0	3
MC3	2	1	3
GC	1	0	1
Total Awards	9	3	12

6 PARTICIPANT DISCUSSION WORKSHOP SESSION

Participant workshops have been held during VisWeek every year since 2008. This workshop combines invited speakers with group discussions and an opportunity for teams to demonstrate their solutions. A participant workshop is being planned for VisWeek 2011 to continue this tradition and provide an opportunity for the teams to interact with one another.

7 THE PATH FORWARD

This VAST Challenge marks the sixth year of the event. This event has consistently attracted significant participation. As stated above, there were more than 671 downloads this year. We have learned a great deal about scenario and data set generation [3, 4]. We know the data sets are being used from email requests, downloads, and citations. Classes and software companies continue to use them and thus they represent a valuable asset as benchmark data sets for the visual analytics community.

Preparing the data sets and running the Challenge are labor-intensive activities. The committee along, with numerous students and staff members, worked not just on the synthesis but also on organizing the reviews, judging, identifying the awards, setting up and running the workshop. The value is clear but the future of the challenges relies upon community support in order to continue.

8 ACKNOWLEDGMENTS

The Committee acknowledges the US Department of Defense for helping support the VAST Challenge. Members of the committee were also supported in part by the National Science Foundation (0947343 and 0947358).

The Committee also wishes to thank Catherine Plaisant and Manas Desai at the University of Maryland; Adam Roberts, Bill Pike, Jean Scholtz, and John Burnette of Pacific Northwest National Laboratory; and all the students of the University of Massachusetts Lowell, especially John Fallon, Patrick Stickney, Heather Byrne, Baochen Sun, Yen-Fu Luo and Loura Costello.

9 REFERENCES

- [1] VAST Challenge: <http://www.cs.umd.edu/hcil/vastchallenge/>
- [2] Visual Analytics Benchmark Repository:
<http://hcil.cs.umd.edu/localphp/hcil/vast/archive/>
- [3] Costello, L., Grinstein, G., Plaisant, C. and Scholtz, J., Advancing User-Centered Evaluation of Visual Analytic Environments through Contests, *Information Visualization* 8 (2009) 230–238
- [4] Whiting, M., Haack, J., and Varley, C. 2008. Creating realistic, scenario-based synthetic data for test and evaluation of information analytics software. In *Proc. of BELIV'08*, ACM, New York, NY.

Methods for Extracting Social Network Data from Chatroom Logs

O. Isaac Osesina^{*a}, John P. McIntire^b, Paul R. Havig^b,
Eric E. Geiselman^b, Cecilia Bartley^a, M. Eduard Tudoreanu^a

^a Dept. of Information Science, University of Arkansas at Little Rock, AR 72204, USA

^b 711th Human Performance Wing / RHCW, US Air Force Research Laboratory, Dayton, OH, USA

ABSTRACT

Identifying social network (SN) links within computer-mediated communication platforms without explicit relations among users poses challenges to researchers. Our research aims to extract SN links in internet chat with multiple users engaging in synchronous overlapping conversations all displayed in a single stream. We approached this problem using three methods which build on previous research. Response-time analysis builds on temporal proximity of chat messages; word context usage builds on keywords analysis and direct addressing which infers links by identifying the intended message recipient from the screen name (nickname) referenced in the message [1]. Our analysis of word usage within the chat stream also provides contexts for the extracted SN links. To test the capability of our methods, we used publicly available data from Internet Relay Chat (IRC), a real-time computer-mediated communication (CMC) tool used by millions of people around the world. The extraction performances of individual methods and their hybrids were assessed relative to a ground truth (determined a priori via manual scoring).

Keywords: Social Network Analysis, Social Network Graph, Temporal Analysis, Conversation Cycle, Computer-Mediated communication, Data Mining

1. INTRODUCTION

1.1 Background

Computer-mediated communication (CMC) is an important communication medium in today's developed society. It permeates virtually all aspects of personal and business life and is used for different activities such as banking, socializing, event organizing, etc.[2]. The persistent nature of many forms of CMC affords unique analysis of historical events and presents organizations with a treasure trove of information for customer analysis, decision support and business intelligence. Although the explosion of online user-generated content through social media and other CMC platforms provides a wealth of information about people (sentiment, interests, etc.), it is not without its limitations. For example, fragmented, agrammatical, and interactionally disjointed communications imposed by computer messaging systems create significant challenges toward automatically extracting SN information from CMC. In particular, cross-turn coherence makes the extraction of social network information more complex; especially within massive archives of internet chat data [3].

This research examines synchronous CMC environments where explicit social network relations among users engaging in multiple (disjointed) conversations within the same textual stream are ill-defined; in other words, it is unclear who is talking to who. An example of such an environment is an internet chat room. We analyzed publicly available chat logs from an Internet Relay Chat (IRC) channel. IRC is a real-time CMC tool used by millions of people around the world [1]. It facilitates direct communication among users within "groups" or "channels". The channels--usually organized by shared concepts, common events, interests, hobbies, etc.--contain sequential conversations arranged by timestamps and displayed via a single textual "stream." Interpretation of Social Network (SN) user relationships within a channel is easily performed by a human reader, but the lack of explicit SN information (e.g. conversation start/stop points and sender/receiver identification) creates challenges for the automatic analysis of social network within internet chat rooms.

^{*}oiosesina@ualr.edu

1.2 Related Works

Analyzing social network information in CMC has generated user and researcher interest for reasons such as understanding the evolution of online communities [4], promoting pro-social behaviors [5], increasing social consciousness/encouraging user participation [6], and improving user interaction/increasing satisfaction [7].

Netscan [8], a CMC analysis tool which reveals insights into the structure and characteristics of online communities, has been extensively studied and used for extracting different types of social network information. Burkhalter and Smith's [5] study of Netscan revealed the use of SN information for member "typification", status, and group comparison. Smith et al. [9] studied the similarities between spatial and physical interactions using Netscan. Also, Krikorian and Kiyomiya [10] developed the newsgroup death model for determining the decline of online communities and detection of cliques based on asynchronous user interactions [11]. Conversation Map [12] analyzed CMC information for revealing member centrality, conversation groups, and citation patterns within Usenet newsgroups. Rosen et al. [13] illustrated the use of semantic analysis to determine groups and organizational patterns within an online educational universe.

Neumann et al. [14,15] used unique characteristics of individuals' typing styles in a chatroom to provide an artistic rendering/visualization of the chat. The goal was to improve social consciousness and personalize interaction. PeopleGarden is a visualization that facilitates the analysis of virtual forums [16]. It uses flowers, petals, and pistil-like circles to represent users, their postings, and response quantification respectively. Conversation threading aids SN analysis by improving the usability, navigability, and understandability of conversations and topics under discussion in a chat room. Smith et al. [9] studied automated "conversational threading" within chat in an attempt to restore the natural turn-taking structure of spoken communication which is often disjointed in chat syntax. Rohall et al. [17] performed similar threading work using email CMC in order to counter the disjointedness that is sometimes present in lengthy email exchanges. Angluin et al. [18] used information about illness outbreaks from the Center for Disease Control to infer the underlying real-world social networks of patients to help identify, track, and contain spreading illness outbreaks.

De Choudhury et al [19] discussed the *inference problem*; that "real" social ties are not directly observable and must be extracted or inferred by observing communication events in social networks. The level of challenge in inferring social networks from CMC depends on the properties of the medium used. In a medium such as email, inferring social networks may be easier since observed events (e.g., Tom emails Jerry) are clearly defined and likely to represent a social network link. However such links are not clearly defined in multiuser chatrooms. Tyler et al [20] and Diesner et al [21] explored methods for analyzing social network within e-mails. Also, Eagle et al [22] analyzed friendship social network from mobile phone logs. Since social links do not already exist in chatroom data, these methods cannot be directly applied to inferring links in our dataset.

Mutton [1] introduced several approaches for extracting/inferring SN links within internet chat, two of which are implemented in the present work: (1) temporal proximity of messages and (2) direct addressing of users. We additionally implement message content via keyword-based similarity to infer whether two individuals are talking about the same topic (are in the same conversation). As discussed above, there are a variety of potential uses and applications for social network data in CMC; our research focuses primarily on inferring/extracting links between interacting users and obtaining an overview of the conversation content.

1.3 Research Overview

We present three independent methods: (1) response-time, (2) word context usage, and (3) direct addressing analysis for extracting SN information within an internet chat room. The response-time analysis extracts SN links based on the temporal proximity of chat messages, the word context usage analysis extracts links based on the usage patterns of keywords within each message, and the direct addressing analysis uses references to intended message recipient to identify SN links. Our research aims to answer two main questions: (1) who is talking to whom?, and (2) what is the context of the conversation?

We approached the first question through the extraction of social network links and the second by identifying relevant keywords associated to each link.

2. EXTRACTING THE SOCIAL NETWORK

2.1 Social Network Analysis

A social network (SN) is a social structure consisting of individuals (nodes) and their interdependencies (links or edges). This information is well-represented by a graph structure which often facilitates the mathematical measurement and visualization of relationships and interactions among the entities. Our analysis aims to infer, detect, or otherwise extract the social network connections (links) between the communicators (nodes) in a chat room. The goal is to present this information as a social graph where each user is represented by the node and their interactions and conversation context is modeled as edges. Extraction and visualization should help to more easily distill valuable information from large volumes of textual communication datasets (e.g. Albinsson, & Morin, [23]). Instead of manually searching through thousands of chat messages to discern who is talking to whom, how often, and what they are talking about, analysts can automatically extract this information using graph analysis methods and visualization tools. This may help to efficiently analyze chat room user relationships such as proximity, community membership, social status, conversation content, and communication dynamics. The present work focuses primarily on the SN extraction methods for chat data, as opposed to the visualization aspects of the data.

2.2 Response-Time Analysis

Response-Time analysis (RTA) is based on the temporal analysis of likely responses to messages; it assumes a normative ideal of alternating turns to determine the existence of interaction between users. It uses the natural delay between one comment and the next that is required for user information processing and the random variations in this pattern caused by off-line distractions, network delays, etc. to estimate the likelihood that messages by other users are responses to a specific preceding message. The distribution pattern of the time between related messages within the chat room is determined by sampling and manually perusing the chat messages. Figure 1 shows the relative frequency of response for the first 1.5 million chat messages in the MusicBrainz chat log [24]. This response-distribution is then used to automatically estimate the likelihood that a message is a response to an earlier one within the chat stream. To facilitate the calculation of response probability, a mathematical distribution similar to the manually obtained response-time distribution is used. For example, the log-logistic function can be substituted for the response time distribution in Figure 1.

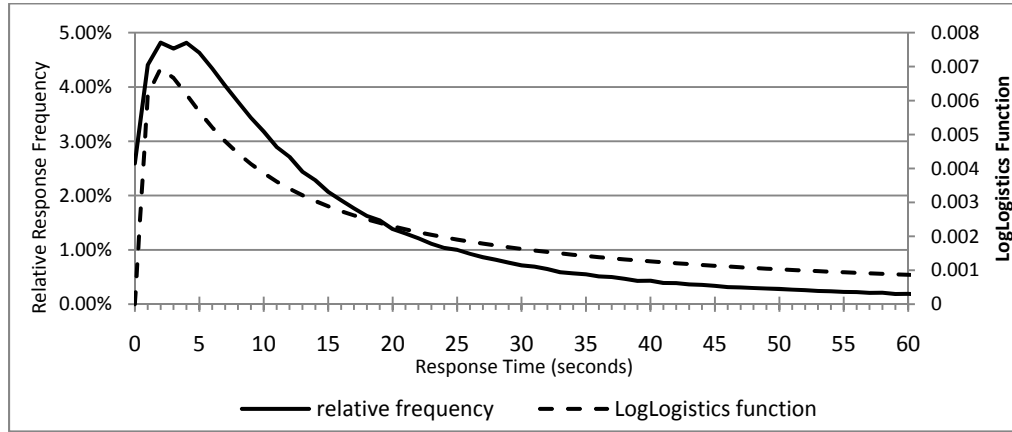


Figure 1: Relative Frequencies of Message Response Times for the first 1.5 million messages in the MusicBrainz dataset compare with the log logistics function where $\alpha=12$ seconds and $\beta=2$

Hence, the probability of response for a message can be calculated using $f(t)$ below. Where t represents the time interval between message and response:

$$f(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{\left(1+(t/\beta)\right)^2} \quad (i)$$

α : is the median response time.

β : reflects the shape of the curve

One identified problem with this approach is that calculating the relationship (message and response) between all messages quickly results in a combinatorial explosion. In order to limit the number of calculations performed to a manageable number, probabilities of response are calculated for only messages that fall within a “conversation cycle”. We introduced the term conversation cycle to describe a message and all its likely responses [24]. It starts with a user’s message and contains the message sequence (from other users) posted before the user’s own next message. For example, in an IRC channel, a conversation cycle for a user starts when they post a message and ends just prior their next message (Figure 2).

3/1/2006 00:23:21	<_rob> one last little thing and then I am taking a break	} Conversation Cycle
3/1/2006 00:24:02	<yeti> later folks	
3/1/2006 00:24:06	<organism> organism has joined #musicbrainz	
3/1/2006 00:24:20	<yeti> yeti has quit	
3/1/2006 00:38:34	<teleMan_> teleMan_ has quit	
3/1/2006 00:38:42	<SenRepa> SenRepa has joined #musicbrainz	
3/1/2006 00:39:57	* SenRepa angrily removes and throws away all "extra" buttons on his keyboard for things like mute	
3/1/2006 00:40:31	<SenRepa> after accidentally opening 30 FireFoxes	
3/1/2006 00:40:52	<JetPower> i dont use those buttons	
3/1/2006 00:41:19	<SenRepa> neither do i	
3/1/2006 00:41:34	<SenRepa> my monitor pressed them	
3/1/2006 00:42:00	<_rob> * _rob goes to take break	

Figure 2: An example of a conversation cycle for user "_rob"

The total level of interaction (weight) among users is determined by the cumulative probabilities from comparable conversation cycles. Assuming that $g_{A \rightarrow B}(i) = f(i)$ represents the probability of user B responding to user A in the i -th conversation cycle, the total level of interaction of user A towards user B ($w_{A \rightarrow B}$) can be calculated as the integral of $g_{A \rightarrow B}$ for all conversation cycles of user A. The interactions of user B directed towards user A can also be calculated in the same manner by reversing the position of A and B. The average of these interactions ($w_{A \leftrightarrow B}$) is a non-directional interaction level between the two users. By iteratively analyzing the interactions for each user (egocentric), this method analyzes the entire social network.

$$\text{Interaction weight } (w_{A \rightarrow B}) = \int_0^{\infty} g(i) di \quad (ii)$$

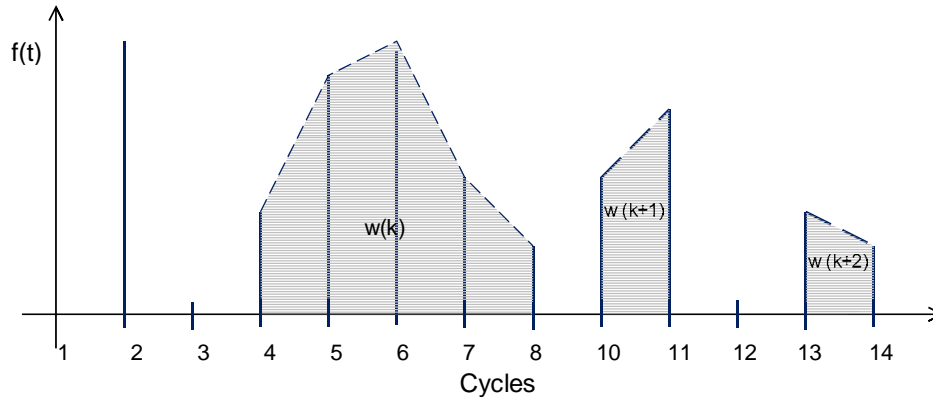


Figure 3: Example of a discontinuous interaction weight function. $w = w(k) + w(k+1) + w(k+2)$

Because this method uses the system time to calculate response probability, it is not directly prone to user generated noise and the conversation language or domain. A limitation of this approach is the difficulty related to manually perusing chat messages in order to determine the empirical response-time distribution. This process is subject to sampling and analyst bias as different samples may yield different distribution patterns and analysts may disagree on any particular message being a response to a reference message. Also, to be effective, this method requires the resolution of the timestamp recorded for each message to be in a similar range as the median response time. For example, if the median time is in the order of seconds, the time stamp must at least have a resolution of one second.

2.3 Word Context Usage Analysis

Word Context Usage Analysis (WCUA) infers SN links among users based on keywords contained in their messages. Keywords are defined as the highest frequency words in the entire user generated chat messages after stop-words removal and are most informative of the conversation content [25,26]. We approached this analysis using two main methods: (1) creating links among users' same/similar keywords and (2) creating links among users' utilizing similar keywords in the same context.

In the first approach, the list of keywords employed by a user is compared to other users' and links are created among users with interacting lists. The relative frequency of usage of each keyword can be used to rank them in the order of importance to the link. Similar frequency of keyword usage between users is more indicative of a link compared to relative skewness in usage. In the second approach, the context within which each keyword is used is determined by creating a link between keywords used in the same chat message. This provides further information for identifying users' involvement (or not) in similar conversations (e.g., users A and B chatting about "bat" the noun and verb respectively). Although this approach can be enhanced by using language and domain resources such as dictionaries and thesaurus, the simple implementation is independent of language and domain. For example, in Figure 4 the single keyword linkage creates a link between the red and green paper figures based on the keyword "music" and the second method would create a link between them based on the keywords "gaga" and "music".

There are two methods of creating SN links through this (second) approach. In both methods, a graph structure can be used for easily managing links (including frequency of joint usage) among keywords. One method creates links among users who used keywords linked in the graph and the other method first clusters the keywords based on the graph modularity and then links users who used linked keywords belonging to the same cluster. By creating natural clusters based on keywords' node properties within the graph, the latter method can help to disambiguate context. For both methods, the lesser frequency of usage of the keywords between the users is assigned as the weight of the link (edge). In addition to providing information for keyword disambiguation and helping to identify correlated or disjointed conversations, WCUA provides information about the possible content of user interactions. A more complex method of creating SN links based on this approach is to create links between users only when they each use keywords in a single message. Although this can help in improving the correlation of conversations, the number of extracted links will be lower.

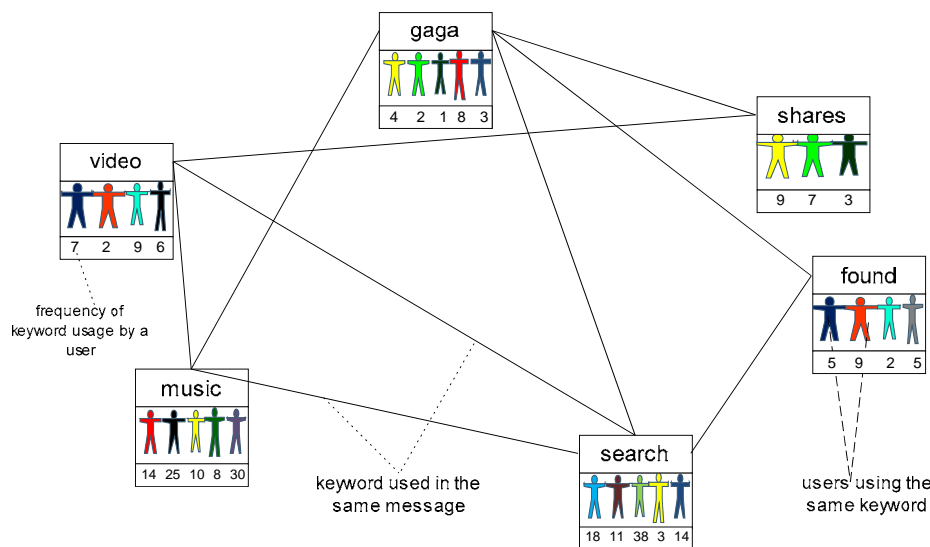


Figure 4: Word Context graph showing links among keywords used in the same message

2.4 Direct Addressing Analysis

Direct Addressing Analysis (DAA), is one of the simplest methods for extracting interaction links among internet chat room users. It infers the intended message recipient by identifying the screen name (nickname) referenced in the message a.k.a. "direct addressing" [1]. Because it is used by the message owner to enhance the clarity of the intended recipient, it provides strong evidence for extracting SN relations in an internet chat room. On the other hand, because

users do not always indicate an intended message recipient this way, the number of links extracted using this method can be expected to be relatively low. Additionally, because it is directly dependent on user input, this method is prone to errors resulting from typographical mistakes and variations of the recipients screen name (user name).

```
<Rob> Is the download complete?
<John> Rob: yes.
```

Figure 5: An example of direct addressing

2.5 Hybrid SN Link Extraction

A hybrid of the three methods may afford the combination of the strengths and unique features of individual methods. For example, the keyword usage analysis can be used to provide context for links extracted using RTA. Also, the independence of the methods to one another increases the confidence of extracted links. Figure 6 depicts an example of a robust SN structure which can be derived by the hybrid SN extraction method. Several methods can be used for combining (to be completed)

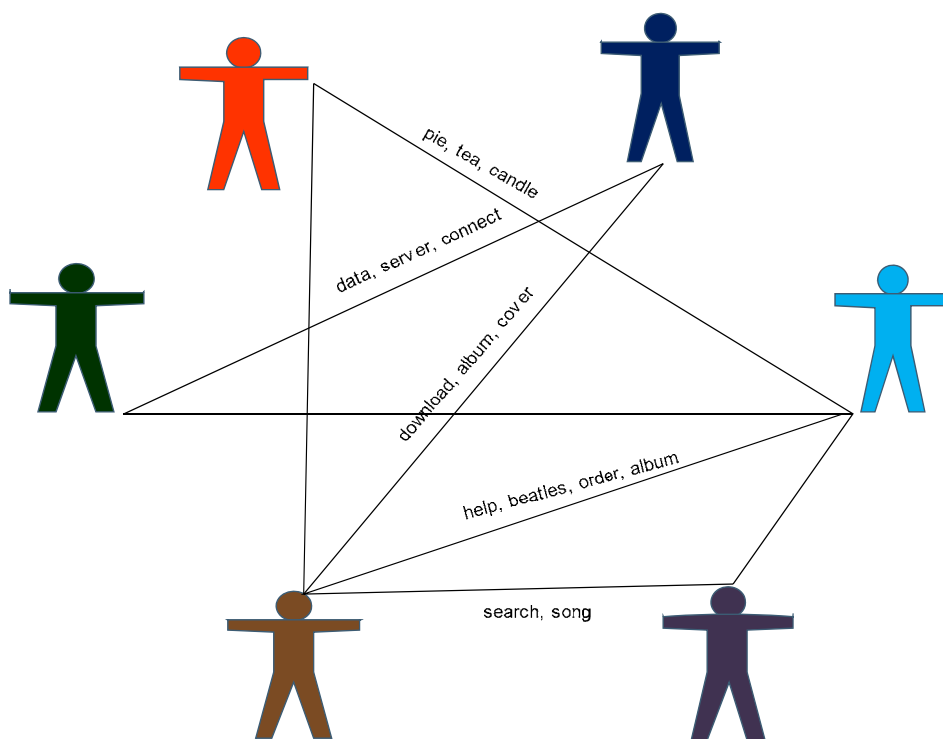


Figure 6: Example of a SN from a Hybrid Method

3. RESEARCH DATASET

3.1 About Dataset

In order to test and evaluate the performance of our methods, we used publicly available internet chat logs from the IRC channel called MusicBrainz (<http://chatlogs.musicbrainz.org/musicbrainz/>). This archive includes recorded conversations spanning several years and is primarily used by music fans to discuss music related topics and for building a music database. This dataset is ideal for our study because it contains interactions among several users engaging in sometimes single and other times concurrent conversations. Additionally, the chat message timestamps are at the same resolution as the median response time for the dataset. This dataset is an ideal test-bed for evaluating the performance of our SN link extraction methods in a real-world application.

We collected the dataset using software designed to crawl the website and obtained the chat content along with their timestamps. Because some of our link extraction methods require managing user identity throughout the entire dataset, and for easy manipulation, we developed pre-processing software to model the data.

3.2 Pre-processing

The software application (written in Java) sequentially processed each the chat messages in order of ascending timestamp. It modeled each message as a software object with defined attributes and methods (e.g. ChatMessage class contains the message date, body, type and userId of the owner and the ChatUser class contains information about user login/log outs, used screen names and keywords). During pre-processing, it managed information about new users, screen name changes, number of users in the chat rooms, etc.

Pre-processing of this chat dataset is similar to processing other IRC channel or general internet chatroom datasets. However, certain functionalities depend on the setup of the chat server; particularly automatic notification messages. For example, the MusicBrainz server uses the construct “<user1> user1 has quit” to indicate that user1 has left the chat room. Parsing of this kind of messages must be customized for each chat server. Our parser was very effective except in a few cases of unexpected user behavior (e.g. users with multiple login names).

4. EXPERIMENT ANALYSIS

4.1 Ground Truth

We assessed the performance of the above described methods relative to a “ground truth” dataset in which the underlying social network in a sample of chat was determined a priori via manual scoring. A dataset spanning 3 continuous days (1 – 3 June, 2006) was chosen for this experiment. The main factor for restricting the number of days in the dataset was to make the process of perusing through the messages and identifying links more manageable. During the manual link extraction, the recipient(s) of each message were identified and the interaction weights between users were obtained from the count of sender and recipient links within the dataset. We conducted two sets of manual scoring using different individual scorers. Between the two groups of scorers, we found a reliability of 0.99 (Cronbach’s alpha) and Pearson correlation of 0.90. Our final set of ground truth data consist only of SN links extracted during both sets of manual scoring (any links with disagreement were discarded). We did not consider the direction of interaction (i.e., Mike talking to Bob is treated the same as Bob talking to Mike) during this experiment, and the final interaction weight is the average of both manually scored sets.

4.2 Correlation

We used Spearman’s rank correlation and Pearson’s standard correlation to measure the effectiveness of our methods for extracting the social network links and for estimating the relative strength of each interaction (link weight) within the network. We calculated the rank correlation between the ranking of each method scores and the ground truth using Spearman’s correlation. In addition, we used scatter plots to visualize the rankings to provide more detailed analysis of the correlations. The standard correlation provides a measure of the effectiveness of determining the strength of individual interactions. We also visualized this correlation using scatter plots for more in depth analysis. For these visualizations, we used normalized interaction scores in order to eliminate the difference in the range of the values.

4.3 Link Retrieval-Recall, Precision and F-measure

From an information retrieval perspective, the performance of each method for extracting links within the dataset [27] was examined, i.e. the performance in identifying the weight of interaction was not considered. *Recall* measures the performance of the method in retrieving all possible links and *precision* measures the accuracy of the retrieved links [28]. F-measure [29] uses both the recall and precision to form a combined score of accuracy for each individual extraction method.

4.4 Experiments Results

The first experiment was performed in order to analyze the performance of the three previously described WCUA implementations. In a second experiment, we compared the performance of the most effective WCUA approach to RTA and DAA. The performance of a hybrid SN link extraction method was compared to that of the individual component extraction methods.

4.5 Comparison of WCUA approaches.

Figure 7 and Figure 8 show the correlation and link retrieval performances of the single, double and modular (double) keywords WCUA approaches. Although the performances of the three approaches are relatively close, the single keyword approach outperformed the two other approaches (apparent in both figures 7 and 8). Contrary to expectations, the modularity (double) keywords did not outperform other approaches. This performance decrement may be attributed to the reduction in the number of keywords considered in this approach (several words occurring alone in a cluster are ignored). An alternative approach to clustering words used in the same context(s) may result in better performance.

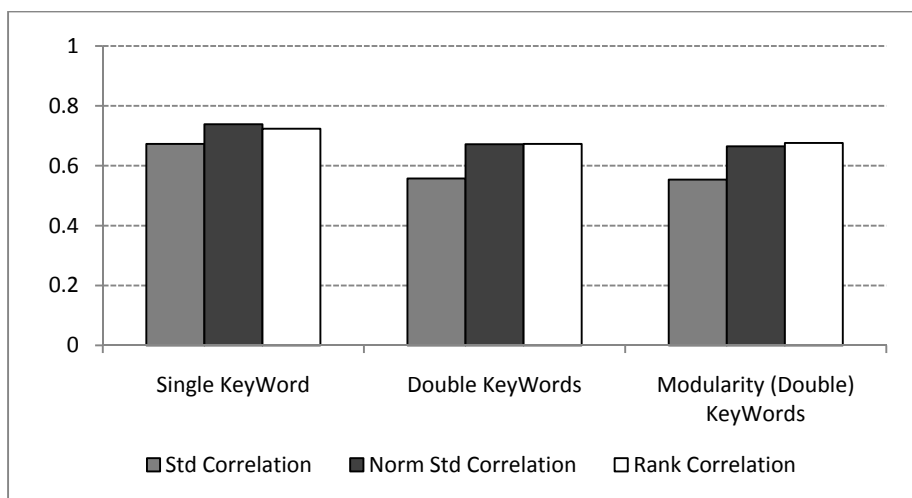


Figure 7: WCUA approaches Correlation Performances

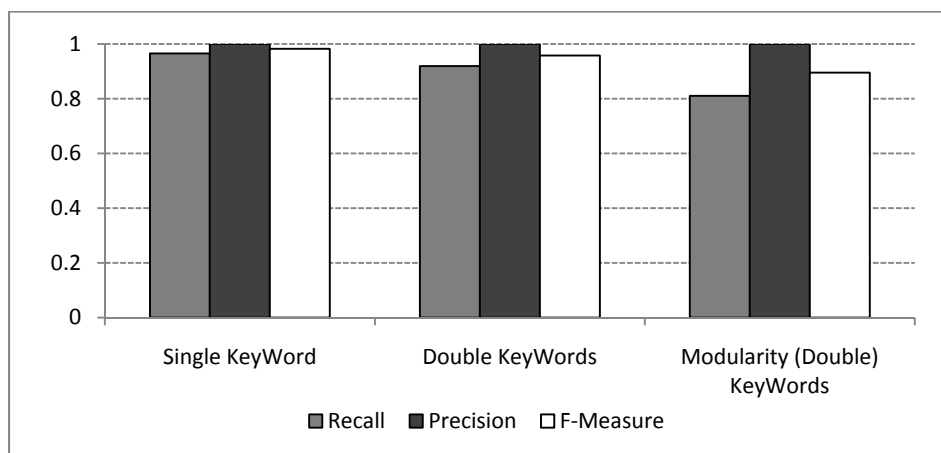


Figure 8: WCUA approaches Link Retrieval Performance

4.6 Comparison of the 3 SN Links methods

By comparing the three SN link extraction methods, we choose the single keyword method because of its superior performance. As shown in Figure 9, the response-time analysis exhibited better performance than the direct addressing and word context usage analyses. As expected, the direct addressing method had the worst correlation performance due to its infrequent use during chat communications.

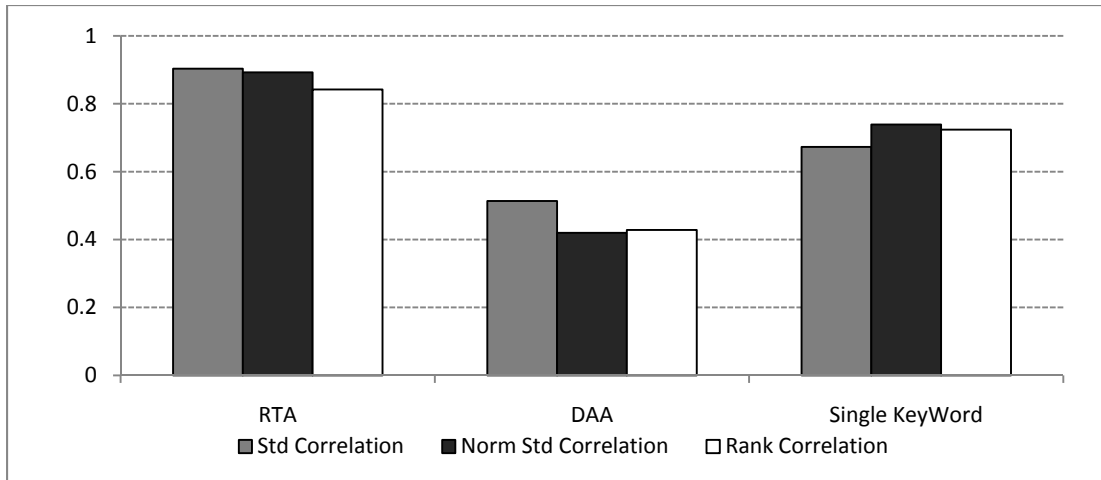


Figure 9: SN Link methods Correlation Performances

Figure 10 shows the link retrieval performance of the three approaches. Although the direct addressing approach had an impressive precision, its ability to extract links is limited by the same reason it has the worst link strength prediction performance. Also, the WCUA outperformed the other two methods in the number of accurate links retrieved.

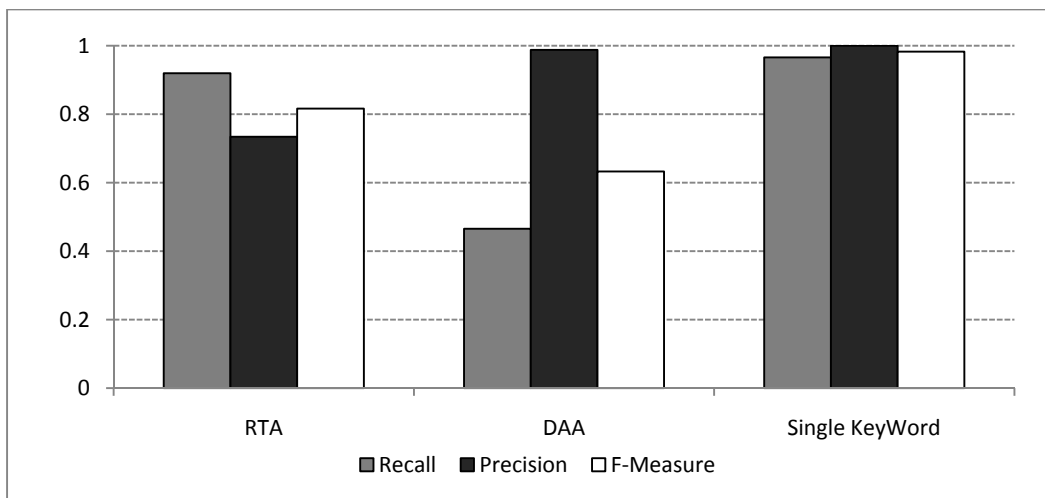


Figure 10: SN Link methods Link Retrieval Performance

4.7 Hybrid SN Link Extraction

A hybrid SN link extraction system that includes a combination of different extraction methods potentially offers better performance. Given that the three methods described above are independent of one another, i.e. their probability of correctly extracting a link is statistically independent of any other; the probability of correctly extracting a link with their ensemble is relatively higher than for the individual methods. Assuming that X, Y, and Z denote identical link extracted using different methods, the probability of correctly extracting the link using a combination of the methods $P(X \otimes Y \otimes Z) = P(X) \cdot P(Y) \cdot P(Z)$. One way of combining these methods is by assigning weights representing their contribution to the hybrid extraction. The value of the weights can be chosen randomly, intuitively or statistically.

In this research, the weight of each method was determined by the normalized values of their f-measure. In other words, the contribution of each method to the hybrid SN links strength, s_{hybrid} was calculated as the product of their normalized link strength and f-measure. The final s_{hybrid} was then calculated as the sum of these contributions i.e. $s_{\text{hybrid}} = \sum (w * s)_{\text{SN}}$

link extraction method, where weight, w is the f-measure and s is the link strength. Using this method we obtained weights of 0.39, 0.36 and 0.25 for RTA, WUCA (single-keyword modularity) and DAA, respectively. Figure 11 and Figure 12 below show the performance (link retrieval and strength correlation respectively) of the hybrid extraction method compared with the individual methods. As seen in the figure, the hybrid method performed better than its individual component method in link retrieval and rank correlation (the two measures considered in the weights calculation) by 1% - 4%.

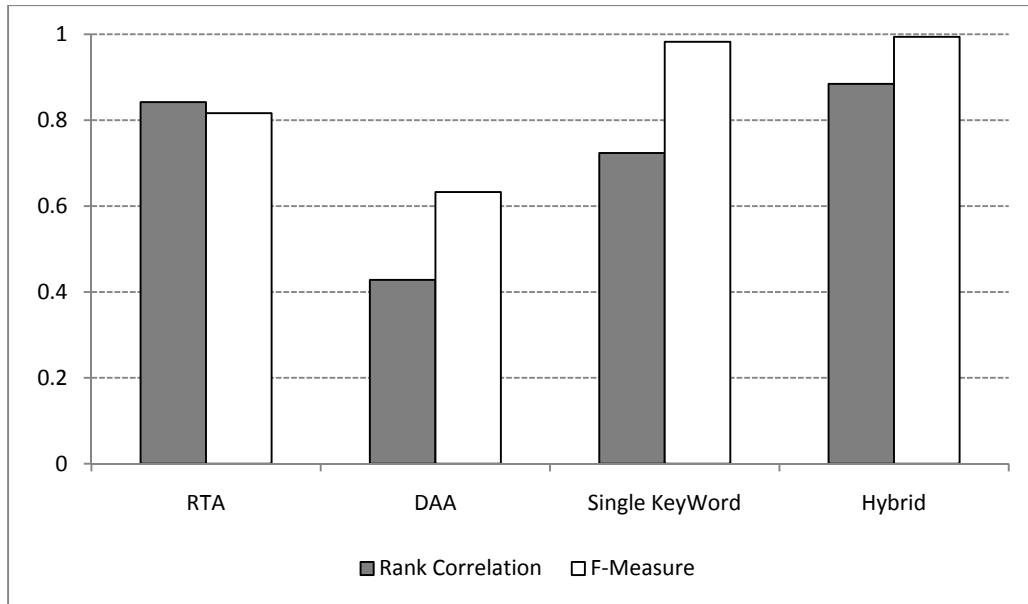


Figure 11: Comparison of Hybrid SN Link Extraction performance to its component methods

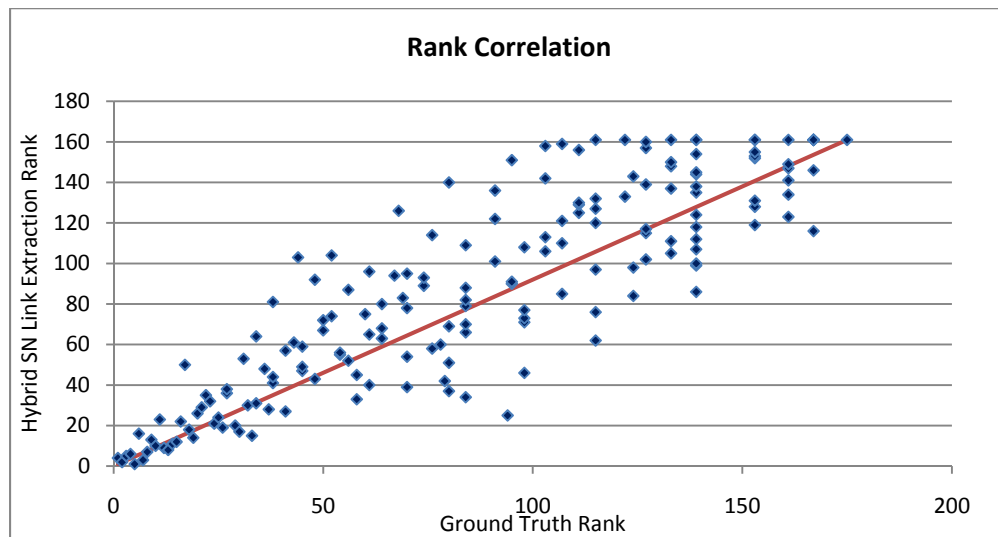


Figure 12: Rank Correlation of the Hybrid SN Link Extraction Method

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

We described three independent methods: (1) response-time, (2) word context usage and (3) direct addressing analyses for extracting SN links within chat room conversations. We also presented methods for combining the different methods into a hybrid system. The performance of these methods was evaluated using publicly available real-world chat room data. A ground truth dataset based on a priori manually scored SN links were used for assessing the accuracy. We evaluated the performances based on interaction weights (correlation) and amount of links extracted (link retrieval).

The WCUA analysis revealed that the single keyword context usage has a slightly better performance than the other two WCUA approaches. Furthermore, it has the best performance in retrieving links although the response-time analysis demonstrated better performance for assigning link strength (highest correlation to ground truth). Contrary to expectations, the modularity double keywords approach did not perform better than the other methods. Generally, both the RTA and WCUA demonstrated high performance for assigning link strength 67%-84% rank correlation and retrieving links more than 80% (f-measure). The direct addressing method showed high precision but low recall (47%) because it was unable to extract links in messages where the addressee was not explicitly mentioned in the message.

5.2 Future Work

Follow-on research will include exploring the extraction of links using contexts containing more than two keywords. Also, other methods for clustering the keywords would be investigated. Research into the use of the keyword context graph for determining cliques; conversation trends, etc. may provide insight into the dynamics of the social network. In addition, other methods for effectively combining the methods in a hybrid system should be researched. Finally, further performance analysis on chat datasets comprised of different languages and domains and alternative forms of CMC should be investigated.

6. ACKNOWLEDGEMENT

We wish to thank the Tec^Edge Summer at the Edge (SATE) program for its invaluable support, particularly that provided by Dr. Robert Williams of the US Air Force Research Laboratory, Dayton.

References

- [1] Mutton, P., "Inferring and Visualizing Social Networks on Internet Relay Chat," Proceedings of the Information Visualisation, Eighth International Conference, 35-43 (2004)
- [2] Baron, N. S., [Language of the Internet], Stanford: CSLI Publications, 59-127 (2003).
- [3] Herring, S., "Interactional Coherence in CMC," Proceedings of the Thirty-Second Annual Hawaii International Conference on Systems Sciences (1999)
- [4] Krikorian, D. and Ludwig, G., "Groupscape: Data mining tools for online communication networks," 22nd Annual Sunbelt Social Network Conference (2002)
- [5] Burkhalter, B. and Smith, M., [Inhabitants' uses and reactions to Usenet social accounting data], Springer-Verlang (2004).
- [6] Kelly, S. U., Sung, C., and Farnham, S., "Designing for Improved Social Responsibility, User Participation and Content in On-Line Communities," Proceedings of CHI (2002)
- [7] Viégas, F. B., "Revealing individual and collective pasts: Visualizations of online social archives," September 2005.
- [8] Smith, M., [Invisible crowds in cyberspace: Measuring and mapping the social structure of USENET], Routledge Press (1999).
- [9] Smith, M., Farnham, S., and Drucker, S., "The social life of small graphical chat spaces," ACM SIG CHI (2000)
- [10] Krikorian, D. and Kiyomiya, T., [Bona fide groups as self-organizing systems: Applications to electronic newsgroups], Lawrence Erlbaum, 335-365 (2002).
- [11] Krikorian, D. and Ludwig, G., "Advances in network analysis: Over-time visualization, dual-mode relations, and

- clique detection methods," 23rd Annual Sunbelt Social Network Conference (2003)
- [12] Sack, W., "Discourse Diagrams: Interface Design for Very Large Scale Conversations," Proceedings of HICSS (2000)
 - [13] Rosen, D., Woelfel, J., Krikorian, D., and Barnett, G. A., "Procedures for Analyses of Online Communities," Journal of Computer-Mediated Communication 8(4) (2003)
 - [14] Neumann, P., Tat, A., Zuk, T., and Carpendale, S., "Personalizing Typed Text Through Visualization," In Proc. Compendium of InfoVis, 138-139 (2006)
 - [15] Neumann, P., Tat, A., Zuk, T., and Carpendale, S., "KeyStrokes: Personalizing Typed Text with Visualization," Eurographics/ IEEE-VGTC Symposium on Visualization, 43-50 (2007)
 - [16] Xiong, R. and Donath, J., "PeopleGarden: Creating data portraits for users," Proceedings of UIST (1999)
 - [17] Rohall, S. L., Gruen, D., Moody, P., and Kellerman, S., "Email visualizations to aid communications," Proceedings of the IEEE Symposium on Information Visualization (2001)
 - [18] Angluin, D., Aspnes, J., and Reyzin, L., "Inferring social networks from outbreaks," 21st International Conference, ALT 2010, 104-118 (2010)
 - [19] Choudhury, M. D., Hofman, J., Mason, W., and Watts, D. J., "Inferring Relevant Social Networks from Interpersonal Communication," Proceedings of the 19th International Conference on World Wide Web, 301-310 (2010)
 - [20] Tyler, J. R., Wilkinson, D. M., and Huberman, B. A., "Email as spectroscopy: Automated discovery of community structure within organizations," Communities and Technologies, 81-96 (2003)
 - [21] Diesner, J., Frantz, T. L., and Carley, K. M., "Communication Networks from the Enron Email Corpus "It's Always About the People. Enron is no Different", " Computational and Mathematical Organization Theory 11(3), 201-208 (2005)
 - [22] Eagle, N., Pentland, A. (., and Lazer, D., "Inferring friendship network structure by using mobile phone data," Proceedings of the National Academy of Sciences 106, 15274-15278 (2009)
 - [23] Albinsson, P.A. and Morin, M., "Visual Exploration of Communication in Command and Control," , 141-146 (2002)
 - [24] McIntire, J., Osesina, O. I., and Craft, M., "Development of Visualizations for Social Network Analysis of Chatroom Text," International Symposium on Collaborative Technologies and Systems (2011)
 - [25] Jones, K. S., "A statistical interpretation of term specificity and its application in retrieval," Journal of Documentation 28, 11-21 (1972)
 - [26] Dalianis, H. et al., [Porting and evaluation of automatic summarization, 107-121 (2004).
 - [27] Singhal, A., "Modern Information Retrieval: A Brief Overview," Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24(4), 35-42 (2001)
 - [28] Cleverdon, C., "The Cranfield tests on index language devices," Aslib Proceedings 19, 173-192 (1967)
 - [29] Kandefer, M. and Shapiro, S., "An F-Measure for Context-based Information Retrieval," 9th International Symposium on Logical Formalizations of Commonsense Reasoning, 79-84 (2009)
 - [30] Smith, M., Cadiz, J. J., and Burkhalter, B., "Conversation trees and threaded chats," Proceedings of the 2000 ACM conference on Computer supported cooperative work, 97-105 (2000)
 - [31] Krikorian, D., "The newsgroup death model: Internet groups as self-organizing," Progress in communication sciences 18 (2002)

Why Social Network Analysis is Important to Air Force Applications

Paul R. Havig*, John P. McIntire, Eric Geiselman, and Fairul Mohd-Zaid

The Air Force Research Laboratory 711HPW/RHCV
2255 H. Street Bldg 248, WPAFB, OH USA 45433-7022

ABSTRACT

Social network analysis is a powerful tool used to help analysts discover relationships amongst groups of people as well as individuals. It is the mathematics behind such social networks as Facebook and MySpace. These networks alone cause a huge amount of data to be generated and the issue is only compounded once one adds in other electronic media such as e-mails and twitter. In this paper we outline the basics of social network analysis and how it may be used in current and future Air Force applications.

Keywords: Social network analysis, social network, applications

1. INTRODUCTION

In recent years there has been a rise of various “social” networks. Facebook® is the most well-known of them, originally started for college students to maintain connections with each other and communicate more effectively to groups of friends. This network also allows people to update various “status” modes they are in such as, dating, single, etc. Finally, it has also served as a sounding board for users to tell all their friends about their day. This is often in a form more like a news bullet: “Having a bad day” or “Am bored who wants to come over and have a party?” More recently, much to original users’ chagrin, it has become a hot bed of companies attempting to use this new medium to reach out to customers and be “liked” by them. Companies have swarmed to this networking capability as it brings marketing to a whole new level as users give much information on their pages that can then be used to direct marketing campaigns and even develop new products (or cease ones that are not working well). Culturally the use of Facebook® has been linked to gatherings beyond a simple party to things such as; flash mobs, public protests, as well as protests against products that remain only a “virtual” protest. Social network platforms played a key role in the Arab Spring political uprisings that occurred in 2011, allowing protestors and revolting militias to document the events as they occurred and to communicate effectively in organizing and fighting against their respective governments, all in near-real time.

However, this type of social network is not only beneficial for keeping up with friends and/or spreading marketing news. For those whom like to maintain and grow business relationships there is LinkedIn® which could be considered a business social network. Users of this social network work in the same way as Facebook®, friending people they know but the idea for LinkedIn® is to connect to past, present, and even future business associates. As a business social network it serves to maintain a sense of business community when associates are geographically dispersed. Further, many professional organizations have LinkedIn® profiles so that they may more effectively provide information to their member base. This includes information about upcoming calls for papers, conferences, and even hosting discussion sections on topics of interests. For those who remember the internet before there were so many pictures this was often done via Usenet groups (e.g., comp.lang.pascal which was a Usenet group to discuss programming in the Pascal language).

In general a social network is any linking of people and information regarding how they are connected. This being the case one may ask why the Air Force would have any interest in social networks and more specifically social network analysis¹. We will endeavor in this paper to outline what it means to analyze a network as well as showing historic use case for doing such. Finally, we will discuss the reasons we believe analyzing social networks is a current and future need for Air Force applications.

2. WHAT IS SOCIAL NETWORK ANALYSIS?

The traditional definition of social network analysis (SNA) is the representation of social networks as diagrams of arcs and nodes and applying mathematical techniques associated with networks, such as Graph Theory, to analyze the structure of the social networks and relationships between the entities.¹ At first blush most social network visualizations look very much like a tangle of nodes and links; this is especially true as the networks get very large and complex. It is in this situation in which many of the ideas of graph theory² come into play to help organize and clarify the data so a more digestible visualization can be constructed. However, the main point of SNA is to better understand the relationships that are found in the analysis.

3. HISTORY

One of the first discussions of SNA was by Moreno³ who introduced the idea of social structures as network diagrams and coined it “sociometry” when he did a study on an epidemic of runaways at the Hudson School for girls in upstate New York. The idea of representing social groups as networks of relation was not firmly established until the works of Roethlisberger and Dickson⁴ and Warner and Lunt⁵ as well as Homans⁶ development of the matrix based analysis.

Work in this area continued for decades with key methodological landmarks in the development of SNA covered in texts by Burt⁷; Freeman, White and Romney⁸; Wasserman and Faust⁹; Wasserman, and Galaskiewicz¹⁰; Scott¹; Carrington, Scott, and Wasserman¹¹; and the recently published Handbook of Social Network Analysis by Scott and Carrington¹².

Another front of developments took a more applied approach by studying the effects of different communication structures on the speed and accuracy with which a group solve problems which lead to works by de Sola Pool and Kochen¹³, which went unpublished for over twenty years, and by Milgram¹⁴ on the “small world” problem. This problem is also known as the “six degrees of separation” issue in which most people can be linked to anyone else in the world by as few as six links.

4. TECHNIQUES

There are two main techniques used for analyzing social networks; mathematical and visual. The aim of this paper is not to exhaustively review the techniques instead it is a very brief outline with pointers to the literature for those interested. Mathematically, graph theory is the predominant approach and provides the core of formal social network analysis. The benefit of applying graph theory is that it affords a pure mathematical technique to be applied that does not need to consider the semantic content (e.g., who the players are in the network). Further, when networks are presented in matrix form, graph theory can be applied directly without the need for construction of an actual visual representation which is an advantage for large scale data. This is especially important for getting a base understanding before attempting to dig in deeper and visualize relationships.

Another important feature of SNA is to formally understand not only who is connected to whom but also the degrees of connectedness. Often this can be a simple count as given in the six degrees of separation problem mentioned above.

Often however, more information is needed to glean a better understanding of the hierarchy of a network. In this case, centrality measures have typically been used to measure power and influence of an individual node by investigating cliques and clusters.

Another matrix-based approach, referred to as ‘block models’, focuses on the characteristics of social positions, roles, and categories instead of the properties of the individuals. Block models are rigorous methods of matrix clustering that organize networks into hierarchical positions which are central to the role-theoretic concerns of sociology as defined by Nadel¹⁵.

While the use of mathematical techniques can be quite elegant and useful they can also prove to be quite daunting to the naïve and math phobic. Thus many researchers also provide visualizations of the networks to help the less math adept understand either the network itself or the results of a mathematical analysis. UCINET is one of the most notable software packages used to implement graph theoretical constructs and has been extended to offer an intuitive and efficient way of undertaking network analyses through graphical approaches. Also PAJEK, which is included as a sub-program within UCINET, offers a way of handling large-scale data sets and using visual methods of representation. These are just two of many information visualization packages that are available to researchers performing SNA.

In our own research we have begun to investigate the human factors and benefits of using different visualizations to look at the same data. We have conducted some preliminary human factors studies on the effectiveness of various network visualization techniques. We pitted the traditional sociogram node-link visualization method against an alternative method using matrix layouts, and asked participants to conduct a variety of network tasks such as identifying strongly connected nodes, identifying whether two target nodes are connected and how, and identifying paths between nodes. As expected, the matrix visualization method resulted in faster response times with no decrease in accuracy on several of the tasks. For simpler tasks, matrix-based visualizations are generally superior for identifying connections and relationships, but for complex path-tracing type-tasks, a network diagram is sometimes a necessity (as these tasks are barely possible with the matrix format). Figure 1 below is an example from a paper at this year’s SPIE conference¹⁶.

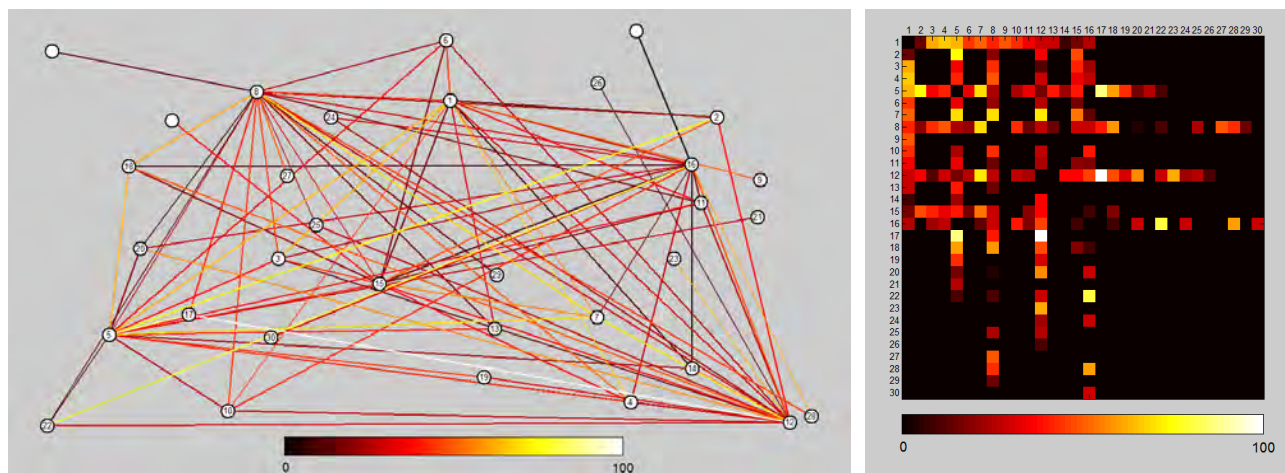


Figure 1. *Left:* The node-link network diagram visualization. *Right:* The adjacency matrix heatmap visualization. Both visualizations represent the same underlying network data. Connection (link) strength is represented by color/intensity (redundantly coded), with corresponding legends shown below both visualizations.

5. ADVANCEMENTS

Recently, there has been a growth of interest by physicists which contributed to advancement in understanding network dynamics and change over time which has been developed weakly by sociologists. Wasserman and Pattinson^{17,18} as well as Robins, Pattinson, and Wasserman¹⁹ developed the 'exponential random graph models' which define a probability distribution on the set of all networks that can be constructed on a given set of points using specific parameter vectors. The benefit of this method is that it can then be used to produce Monte Carlo estimations for statistical tests of significance.

An additional advancement is the use of 'agent-based modeling' which has also been implemented into SNA to study the dynamic global transformation of the network structure over time instead of the traditional static approach. Snijders^{20,21}, following his work with his colleague, Snijders and van Duijn²², developed an approach that sees incremental changes of individual action to the changing network structure and is currently working on making connections with Wasserman's exponential graph models. Snijders' SIENA program implements the overall approach for easy use.

Although SNA itself is not a new area of research, there has been little or no theoretical establishment on the subject. Recently Borgatti, Mehra, Brass, & Labianca²³ proposed an interdisciplinary effort with researchers across other fields of natural sciences to help establish social network theory following the trend of many information visualization teams becoming very interdisciplinary by necessity. Part of the explanation for the current exploding interest in social networks is two-fold: (1) social network datasets are generally large, and so gathering the data was difficult without the digital support afforded by modern communications technology; and (2) analyzing the datasets was very difficult without the computational support afforded by modern computers.

In our research group we propose that the many layers of social networking and electronic communication can conceptually be visualized as the weave of a complex fabric. Due to the ubiquitous participation in this network by the population, the surface of the fabric is smooth and, because of the relatively small degrees of separation between individuals, the surface is also regular. Granovetter²⁴ argued that strongly tied entities provide less meaningful information since the information they provide are likely shared amongst those connected to them. Weak ties on the other hand can easily be unconnected to the rest of the network and more likely to be sources of novel information. An individual who is purposely not participating at any level within the composite network will effectively be conspicuous by their absence. They are represented by a hole or an unusual irregularity in the network weave. Of course there are innocuous reasons for individuals to choose to not participate in these common communication affordances, but it can be confidently stated that nonparticipation will become increasingly rare unless an individual is purposefully attempting to remain undetectable. These holes or dark matter, when detected, may prove to be a reliable indication of suspicious behavior. Missing information may in fact be quite meaningful. The structure of the network may also lend itself to the ability to indicate participants who may be "close" to the missing individual but not electronically "connected" to them. They will be active around the edges of the dark matter. Similarly, large areas of irregularity or patchy holes in the network surface may be indicative of and organized group purposefully attempting to avoid detection and traceability.

To add to this idea of network dark matter being suspicious, several years ago the CIA realized bin Laden was using couriers to issue orders/commands, and he was "famously insistent" that no phones, computers, or other "wired" electronic equipment be used near him. The million dollar mansion compound to which a suspected courier was tracked revealed something rather odd: the complex had no internet or phone, and almost no traffic in or out. This set off alarm bells for U.S. intelligence analysts, ultimately leading to his killing by U.S. military forces^{24,25}.

6. APPLICATIONS TO AIR FORCE

The widespread use of SNA is not ubiquitous to Air Force applications but is coming on strong. In a previous paper²⁶ we have discussed the term network-centric warfare (NCW) and how it has become quite a buzz worthy term. What we discussed were the tenets of NCW as outlined by Alberts²⁷:

- 1) A robustly networked force improves information sharing.
- 2) Information sharing and collaboration enhance the quality of information and shared situation awareness.
- 3) Shared situation awareness enables self-synchronization.
- 4) These, in turn dramatically increase mission effectiveness.

In this paper we had actually, unknown to us, started looking at NCW in terms of SNA. In this case the example was looking at commander and soldier level communication and is ripe for SNA. The example went as follows. We could have multiple groups with the same or a similar set of goals (e.g., destroy buildings A, B, then C) but with varying levels of network interconnectivity. The levels of “network-centrality” could be varied in each group by permitting or restricting communication between nodes within the network or by restricting information flow to a single direction through some nodes. For example, in Figure 2, group 1 might only have one-way communication with their commander, who hands down the orders that are dutifully followed with little or no communication thereafter. In this case, there are few connections between nodes and information is primarily flowing in only one direction.

Likewise, a second group might allow two-way communication between a single foot soldier and his or her commander. In this instance, the speed and efficacy of communication between the commander and the foot-soldiers is limited by the go-between, who relays the messages of the commander to the remaining soldiers (and vice-versa). Information can easily flow to and from the central node (the go-to soldier) but again, the limited number of connections between nodes could hamper the flow of communication as shown in Figure 3.

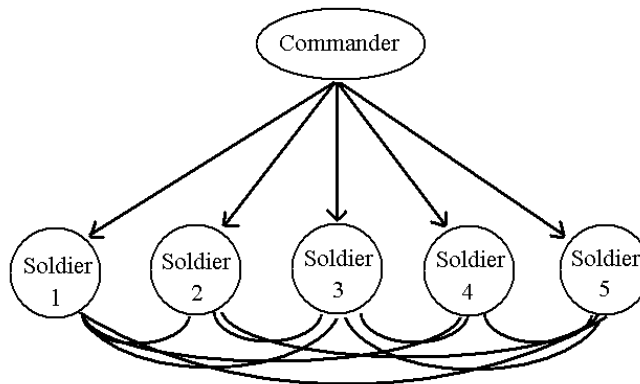


Figure 2 - Group 1. Information flow is primarily one-way, from the commander to the group, although there is unrestrained connectivity between the group members themselves. Notice that there are 5 one-way links and 10 two-way links.

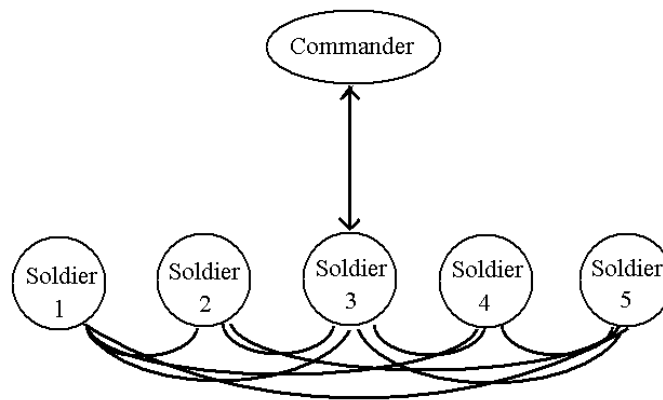


Figure 3 - Group 2. Information between the commander and the group flows (two-way) through a single group member (Soldier 3). Again, there is unrestrained connectivity between members of the group. Notice that there are a total of 11 two-way links.

A third group might allow for simultaneous communication between all nodes of the network, i.e., all soldiers can communicate directly with each other and with the commander. In this case, there is high connectivity and no limitation regarding the flow of information between nodes of the network as shown in Figure 4.

While setting up these types of networks may not be new and indeed may have been tested at various levels from simple to communication to information transfer, we posit to use them only as independent variables that are manipulated to assess higher-level functioning of group behavior. By assessing the performance (e.g., situation awareness, goal effectiveness, decision making, etc.) of each of these groups, one might be able to quantify the effects of numbers and types (i.e., one-way or two-way) of connections between nodes. Thus, we could directly test the advantages gained by using NCW, especially tenet number one. More independent variables that might be beneficially studied using such techniques are: varying the forms of communication that link the nodes (voice + video communications, voice communications, instant-messaging, etc.). Additional dependent variables which may be beneficial to study are: the speed of information flow through the network, the quality of information flow (message degradation), etc. One might even argue that parts of tenet two could be tested as a certain message could be passed up (or down) the chain of command and then tested as to its accuracy thus looking at the effect of quality of information due to collaboration.

In a more recent article about the importance of SNA General McChrystal has some very timely quotes in terms of the importance of understanding networks in fact his paper in *Foreign Policy* is entitled "It takes a network"²⁸.

"Over time, it became increasingly clear -- often from intercepted communications or the accounts of insurgents we had captured -- that our enemy was a constellation of fighters organized not by rank but on the basis of relationships and acquaintances, reputation and fame."

"A true network starts with robust communications connectivity, but also leverages physical and cultural proximity, shared purpose, established decision-making processes, personal relationships, and trust. Ultimately, a network is defined by how well it allows its members to see, decide, and effectively act."

"It takes a network to fight a network."

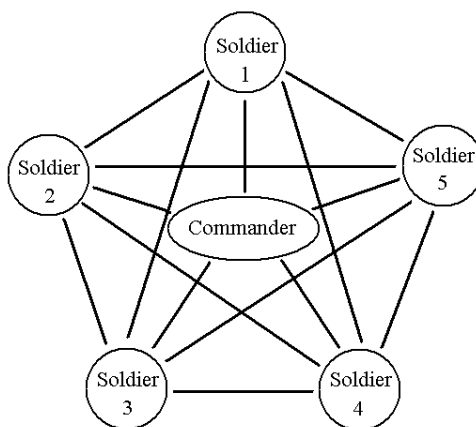


Figure 4 - Group 3. Information flows unrestrained through the entire network, which is fully inter-connected. In this group, there are 15 two-way links.

It is important to note that SNA has been attributed to the finding of both Saddam Hussein as well as Ossama Bin Laden. For Saddam Hussein, once Iraq was invaded by coalition forces, Saddam's official cronies scattered. The famous "deck of cards" put out by the U.S. Defense Intelligence Agency listed the 55 most wanted individuals in Iraq that were mostly political or military cronies of Saddam. But this social structure collapsed amidst the chaos of the coalition invasion, and so these were not the people Saddam would seek refuge from; instead, he would turn to relatives and friends of which the intelligence services knew relatively little about. This obviously complicated the manhunt.

Intelligence and military services on the ground in Iraq started to piece together a social network diagram of everyone they captured or wanted to capture. No hierarchical, top-down graph was evident (since this type of organizational structure was splintered along with the political system); what emerged was instead a convoluted web-like network diagram indicating social connections, with The Butcher of Baghdad at the center.

Investigating this new type of organizational structure eventually led to an interesting and surprising connection to someone not on the deck of cards of high-value targets, and someone not on any intelligence service's radar. It led to an obscure bodyguard whose name was not known, and whose picture led to the temporary moniker "The Fat Man." Following the leads provided by the network structure led to a fishing buddy of the Fat Man, who revealed a possible safehouse owned by The Fat Man, with a possible "spider hole" where someone could hide...leading to the ultimate capture of Saddam²⁹.

A similar story has been told regarding Ossama Bin Laden and has been visualized in this info-graphic to show a social network visualization of his leadership that specifically calls out the black hole idea, saying "sometimes a network is charted by its missing links."³⁰ The important take away here is that SNA can be used to find people of interest. Also, there has been research on reconstructing social networks using various communication strategies: Using email to determine social structure³¹, Using mobile phone data to determine social structure³², and using communications data to determine social structure (not just email or phone)³³.

Our current research has investigated situations in social networks in which the links between nodes are not necessarily explicit, as they are in email or cell phone systems, simply inferring the network structure can be a difficult/challenging task. For instance, in public chatrooms, it may not be immediately obvious who is talking directly to who, and what are the relationships between users. And even if such links are obvious to a human observer poring over the communication logs, getting computer programs to automatically extract these links may be a problem. In previous works, we have presented some visualization ideas³⁴ and developed some automated algorithms that can help extract the social network

structure in such non-explicit communication systems³⁵. In this latter work, we used several techniques to infer the social network links between nodes, including temporal patterns of human communication, content analysis to see whether there were common topics of conversation, and "direct addressing" behaviors in which people explicitly identify the intended recipient of their message (to avoid confusion in a "noisy" chatroom).

7. CONCLUSIONS

Overall we feel that we are just at the beginning of seeing how SNA may apply to Air Force applications. The examples cited are just the tip of the iceberg as to SNAs usefulness and functionality. We aim to extend our research into the information visualization domain and leverage the mathematics used in SNA to make more useful visualizations for better, quicker, and more actionable decisions.

REFERENCES

- [1] Scott, J. (2007). *Social Network Analysis 2nd ed.*, London, Sage.
- [2] Trudeau, R.J. (1993). *Introduction to Graph Theory*, New York, Dove.
- [3] Moreno, J. L. (1934). *Who Shall Survive?*, New York, Beacon Press.
- [4] Roethlisberger, F. J., and Dickson, W. J. (1939). *Management and the Worker*, Cambridge, Harvard University Press.
- [5] Warner, W. L., and Lunt, P. S. (1941). *The social networks as stochastic processes*, New Haven, Yale University Press.
- [6] Homans, G. C. (1950). *The human group*, London, Routledge.
- [7] Burt, R. S. (1982). *Towards a structural theory of action*. New York: Academic Press.
- [8] Freeman, L. C., White, D. R., & Romney, A. K. (1989). *Research methods in social network analysis*. New Brunswick: Transaction Books.
- [9] Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.
- [10] Wasserman, S., & Galaskiewicz, J. (1994). *Advances in Social Network Analysis*. Beverly Hills: Sage.
- [11] Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press.
- [12] Scott, J., & Carrington, P. J. (2011). *Handbook of social network analysis*. London: Sage.
- [13] de Sola Pool, I., & Kochen, M. (1978). *Contacts and Influence. Social Networks*, 5-51.
- [14] Milgram, S. (1967). The Small World Problem. *Psychology Today* vol 2, 60-67.
- [15] Nadel, S. F. (1965). *The Theory of Social Structure*. Glencoe: Free Press.
- [16] McIntire, J. P., Osesina, O. I., Bartley, C., Tudoreanu, M.E., Havig, P.R., & Geiselman, E. (in press). "Visualizing Weighted Networks: A Performance Comparison of Adjacency Matrices versus Node-link Diagrams." *Proceedings of SPIE 2012*, Baltimore, MD.
- [17] Wasserman, S., & Pattison, P. (1999). Logit models and logistic regressions for social networks: I. An introduction to Markov random graphs and p*. *Psychometrika* 60, 401-426.
- [18] Pattison, P., & Wasserman, S. (1999). Logit models and logistics regressions for social networks: II. Multivariate relationships. *British Journal of Mathematical and Statistical Psychology* 52, 169-193.
- [19] Robins, G., Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social network: III. Valued relations. *Psychometrika* 64, 371-394.
- [20] Snijders, T. A. (2001). The statistical evaluation of social network dynamics. In M. E. Sobel, & M. P. Becker, *Sociological methodology*. Oxford: Basil Blackwell, 361-395.
- [21] Snijders, T. A. (2005). Models for longitudinal network data. In P. J. Carrington, & S. Wasserman, *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press, 215-247.
- [22] Snijders, T. A., & van Duijn, M. A. (1997). Simulation for statistical inference in dynamic network models. In R. Conte, R. Hegselmann, & P. Terna, *Simulating social phenomena*. Berlin: Springer, 493-512.
- [23] Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science* vol 323, 892-895.
- [24] Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology* vol. 78 issue 6, 1360-1380.

- [24] http://www.huffingtonpost.com/2011/05/02/osama-bin-laden-killing_n_856633.html
- [25] http://www.huffingtonpost.com/2011/05/02/osama-bin-laden-dead-one-phone-call_n_856674.html
- [26] Havig, P., Aleva, D., Reis, G., & McIntire, J. (2007). Putting the science back in C2: What do the buzzwords really mean? *In The 12th International Command and Control Research and Technology Symposium: Adapting C2 to the 21st Century*.
- [27] Alberts, D.S., (2002). *Information Age Transformation: Getting to a 21st Century Military*. Washington D.C.: Command & Control Research Program.
- [28] http://www.foreignpolicy.com/articles/2011/02/22/it_takes_a_network?hidecomments=yes
- [29] http://www.slate.com/articles/news_and_politics/searching_for_saddam/2010/02/searching_for_saddam_5.html
- [30] <http://news.muckety.com/2011/05/03/mapping-the-hunt-for-osama-bin-laden/31931>
- [31] Tyler, J. R., Wilkinson, D. M., and Huberman, B. A. (2003), "Email as spectroscopy: Automated discovery of community structure within organizations," *Communities and Technologies*, 81-96.
- [32] Eagle, N., Pentland, A., and Lazer, D. (2009), "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences* 106, 15274-15278.
- [33] Choudhury, M. D., Hofman, J., Mason, W., and Watts, D. J. (2010), "Inferring Relevant Social Networks from Interpersonal Communication," *Proceedings of the 19th International Conference on World Wide Web*, 301-310.
- [34] McIntire, J. P., Osesina, O. I., & Craft, M. (2011). "Development of Visualizations for Social Network Analysis of Chatroom Text." *Proceedings of the Collaborative Technologies & Systems Symposium, CTS 2011, Chicago, IL*, p. 593-597.
- [35] Osesina, O. I., McIntire, J. P., Havig, P. R., Geiselman, E. E., Bartley, C., & Tudoreanu, M. E. (in press). "Methods for Extracting Social Network Data from Chatroom Logs." *Proceedings of SPIE 2012, Baltimore, MD*.

VAST Challenge 2012: Visual Analytics for Big Data

Kristin Cook¹, Georges Grinstein², Mark Whiting¹, Michael Cooper¹, Paul Havig³, Kristen Liggett³,
Bohdan Nebesh⁴, Celeste Lyn Paul⁴

¹ Pacific Northwest National
Laboratory

² University of Massachusetts
Lowell

³ Air Force Research Laboratory

⁴ National Security Agency

ABSTRACT

The 2012 Visual Analytics Science and Technology (VAST) Challenge posed two challenge problems for participants to solve using a combination of visual analytics software and their own analytic reasoning abilities. Challenge 1 (C1) involved visualizing the network health of the fictitious Bank of Money to provide situation awareness and identify emerging trends that could signify network issues. Challenge 2 (C2) involved identifying the issues of concern within a region of the Bank of Money network experiencing operational difficulties utilizing the provided network logs. Participants were asked to analyze the data and provide solutions and explanations for both challenges. The data sets were downloaded by nearly 1100 people by the close of submissions. The VAST Challenge received 40 submissions with participants from 12 different countries, and 14 awards were given.

Keywords: Visual analytics, human information interaction, sense making, evaluation, metrics, contest.

Index Terms: H.5.2 [Information Interfaces & Presentations]: User Interfaces – Evaluation/methodology.

1 INTRODUCTION

The Visual Analytics Science and Technology (VAST) Challenge [1] is a series of contests that aim to advance visual analytics through competition. Started in 2006 and now in its seventh year, the VAST Challenge is designed to help researchers understand how their software would be used in a novel analytic task and determine if their data transformations, visualizations, and interactions would be beneficial for particular analytic tasks. VAST Challenge problems provide researchers with realistic tasks and data sets for evaluating their software, as well as to advance the field in solving more complex problems.

Researchers and software providers have repeatedly used the data sets from throughout the life of the VAST Challenge as benchmarks to demonstrate and test the capabilities of their systems. The ground truth embedded in the data sets has helped researchers evaluate and strengthen the utility of their visualizations.

email: {kris.cook,mark.whiting,michael.cooper}@pnnl.gov,
grinstein@cs.uml.edu,{paul.havig,kristen.liggett}@wpafb.af.
mil,{banebes,clpaul}@nsa.gov

2 SCOPE OF VAST CHALLENGE 2012

The goal of VAST Challenge 2012 was to provide a set of realistic computer network scenarios while pushing the boundaries of big data. The setting of the Challenge is BankWorld, a planet much like Earth, but with a very different geography. For this Challenge, the geography is one large land mass containing several different nation-states. The most important organization on BankWorld is the Bank of Money (BOM). BOM has many offices of various sizes across BankWorld. Each of these offices has many computers active throughout the day. In total, the organization operates about 895,000 machines.

Contestants were asked to focus on two general problems using a visual analytics approach. First, how do you achieve cyber situation awareness across the entire enterprise with such a large number of systems? Second, when something does go awry, can you identify it and the steps needed to resolve the problem?

2.1 Contest Problem

VAST Challenge 2012 consisted of two independent but related challenge tasks set in the fictitious BankWorld. Each challenge consisted of a data set, instructions, and questions to be answered. Unlike previous years, this year's VAST Challenge did not include an overarching Grand Challenge that tied the clues from the individual challenges together.

In previous years, the individual challenge tasks have also been referred to as mini-challenges, and they were originally posed to the participants as mini-challenges. However, given the scope and complexity of handling such gigabytes of data, it seems more appropriate to describe the individual tasks as challenges rather than mini-challenges.

Each challenge had certain constraints and business rules that contestants needed to consider for their analysis. For example, in Challenge 1, BOM offices operate during business hours 7am-6pm in their local time zone. However, the BOM enterprise spans ten time zones. Failure to properly handle the geo-temporal issues prevented proper understanding of the evolving problems across BOM.

2.1.1 Challenge 1: Situation Awareness

Challenge 1 focused directly on cyber situation awareness across BOM. Its overview and task questions read:

The Bank of Money (BOM) Corporate Information Officer (CIO) has assigned you to create a situation awareness visualization of the entire enterprise. This is a considerable challenge, considering that BOM operates from BankWorld's coast to coast. In addition to observing the global situation, he would also would like to be able to detect operational changes outside of the norm. You are provided with two data sets that

span two days of data for BOM. One dataset contains metadata about the bank's network. The second dataset contains periodic status reports from all computing equipment in the BOM enterprise.

MC 1.1 Create a visualization of the health and policy status of the entire Bank of Money enterprise as of 2 pm BMT (BankWorld Mean Time) on February 2. What areas of concern do you observe?

MC 1.2 Use your visualization tools to look at how the network's status changes over time. Highlight up to five potential anomalies in the network and provide a visualization of each. When did each anomaly begin and end? What might be an explanation of each anomaly?

2.1.2 Challenge 2: Operational Forensics

Challenge 2 focused on operational forensics. Its background and task questions were:

During a time period that is NOT overlapping with MC 1, a Region within the Bank of Money is experiencing operational difficulties. This becomes a challenge for the operations staff, particularly as they attempt to deploy their limited number of skilled administrators to address issues occurring in the enterprise. You will be provided with Firewall and IDS logs from one of the BOM networks of approximately 5000 machines. These are very similar to the Firewall and IDS logs you worked on during the VAST 2011 MC 2, and so the tools you used there will come in handy for this mini-challenge (and reuse is encouraged). You will also be provided with a description of the network to guide your investigation.

MC 2.1 Using your visual analytics tools, can you identify what noteworthy events took place for the time period covered in the firewall and IDS logs? Provide screen shots of your visual analytics tools that highlight the five most noteworthy events of security concern, along with explanations of each event.

MC 2.2 What security trend is apparent in the firewall and IDS logs over the course of the two days included here? Illustrate the identified trend with an informative and innovative visualization.

MC 2.3 What do you suspect is (are) the root cause(s) of the events identified in MC 2.1? Understanding that you cannot shut down the corporate network or disconnect it from the internet, what actions should the network administrators take to mitigate the root cause problem(s)?

2.2 Submission Format

Teams were asked to provide a video and a concise process description as to how they arrived at their conclusions and how the various visualizations and tools helped in the analysis.

2.3 Review Process

As in years past, the VAST Challenge Review Committee recruited reviewers from throughout the visualization and analysis communities. Subject matter experts were recruited from the pool of previous reviewers and their social networks.

Including both the visualization community reviewers and the subject matter expert reviewers, a total of 102 reviewers participated, each providing between one and five reviews. This represents a significant increase from the 56 reviewers who participated in 2011. Four to eight external reviewers, including at least one subject matter expert, reviewed each submission. Each reviewer was given the opportunity to recommend submissions for specific awards.

Reviewers were asked to rate the analytic process, the visualizations, the interactions, the clarity of explanation, and the relative novelty of the submission. In addition, reviewers rated the submission in terms of its support for dynamic situation awareness, as well as the identification of specific events of interest in the data. Reviewers provided both ratings and explanatory comments. These comments were as important as the scores in identifying award candidates.

Reviewers were also asked to evaluate the plausibility of the answers provided, rather than the accuracy of the solutions. The datasets used this year were realistically complex. Although there were certain known patterns embedded in the data, the committee recognizes the likelihood that additional patterns exist in the data that were not intended and that could reasonably be considered by the participants to be of significance. Consequently, reviewers were provided with a list of the expected patterns that were embedded in the dataset to support the scenario, but they were also instructed to accept other solutions for which the submission provided well-reasoned supporting evidence.

The VAST Challenge Review Committee held a one-day meeting to determine awards. Prior to the meeting, all of the committee members examined at least nine of the submissions in detail, with five committee members examining all 40 submissions. During the meeting, the committee reviewed and evaluated the award recommendations from the reviewers, taking the totality of the scores and reviewer comments into account. The committee also identified additional appropriate awards.

As in previous years, the awards were not pre-established. Instead, the committee identified awards recognizing the best qualities in the submissions. In addition, this year a few teams were selected to receive honorable mentions. This designation was chosen to recognize entries that demonstrated great promise but were not yet fully realized in their implementation.

3 VAST CHALLENGE 2012 AWARDS

The visualizations required for the two challenges were of substantially different varieties. The geo-spatial and temporal aspects, combined with the enormous number of facilities and machines involved in C1, suggested a different approach than the increasingly odd communication patterns across approximately 5000 machines in C2.

Both C1 and C2 were significant challenges due to the data size and complexity and the difficulties of the tasks specified in each. In general, the challenge participants should be congratulated for their efforts, as reviewers found an abundance of compliments to include in their write-ups. The reviewers and the committee would have liked to have seen even more innovation in the visualizations that would work well for situation awareness. Traditional visualizations (line charts, bar charts, linear and radial graphs, colored geographic areas) were well applied, but future contestants should be encouraged to take more risks in developing new visualizations in support of cyber analytics.



Figure 1: The 2012 award winner for Outstanding Comprehensive Submission was BANKSAFE: A Visual Situation Awareness Tool for Large-Scale Computer Networks (University of Konstanz).

3.1 Comprehensive Award

One group (University of Konstanz) brought the two datasets together to enable situation awareness analytics across both challenges and was recognized for essentially tackling a Grand Challenge problem, even though there was not “official” Grand Challenge in 2012 (Figure 1).

3.2 Challenge 1 Awards

C1 asked for both a static situation awareness snapshot and a dynamic trend-oriented assessment. Two teams (Business Forensics and Charles River Associates) were recognized for visual designs that reviewers felt would work well in an operational setting. Another team (Purdue) submitted an entry that reviewers noted for its outstanding features for integrated analysis and visualization. One team caught the reviewers’ attention by engaging subject matter experts in the design and testing of their toolkit (Middlesex University) and were complimented with a “Subject Matter Experts’ Award.” Other recognitions included Comprehensive Visualization (Stuttgart), Effective Video Presentation (Secure Decisions), Efficient Use of Visualization (City University London), Good Interaction Techniques (General Dynamics C4 Systems), and Good Support for Data Preparation, Analysis and Presentation (MTA Sztaki).

3.3 Challenge 2 Awards

In C2, participants were analyzing the effects of a growing botnet infection across a BOM region, with clues provided in the Firewall and the Intrusion Detection System logs. The geography to consider in this challenge now becomes one of a computer network that was provided in a network diagram with identification of critical systems and business rules of operation.

As for submissions, the committee appreciated that University of Buenos Aires submitted several entries to both Challenges this year. The UBA student team led by Marcos Wolff received an award for Effective Use of Commercial Software, providing a clear and effective identification of trends occurring in the data. Other notable awards included a Good Adaptation of Graph Analysis Techniques submitted by the Chinese Academy of Sciences, Central Michigan University, and Northwestern University team. Two honorable mentions were awarded in C2. Virginia Tech provided an entry illustrating Good Use of Coordinated Displays, and Central South University (China) provided an Interesting Use of Radial Visualization Techniques.

4 DISCUSSION

4.1 Participation

VAST Challenge 2012 received 40 submissions across the two challenges. Table 1 compares the number of submissions over the life of the VAST Challenge.

Submissions	2006	2007	2008	2009	2010	2011	2012
<i>Challenge 1</i>	-	-	22	22	14	30	27
<i>Challenge 2</i>	-	-	13	17	22	8	13
<i>Challenge 3</i>	-	-	12	5	17	13	-
<i>Challenge 4</i>	-	-	20	-	-	-	-
<i>Grand Challenge</i>	6	7	6	5	5	5	-
Total	6	7	73	49	58	56	40

Table 1: Summary of VAST Challenge submissions by year

Given that this year’s challenge involved only two different challenge tasks, as compared to previous years with more available tasks, the number of entries was particularly impressive. The 27 entries received for C1 is the second greatest number of entries received for any of the challenge tasks over the history of the VAST Challenge.

Again this year, the number of dataset downloads has increased. There were 703 unique downloads of the C1 data and 383 downloads of the C2 data, for a total of 1086 unique downloads by the submission closing date. Even accounting for the differences introduced by a new downloading scheme that permits downloads by individual challenge task, this still represents a substantial increase from the 671 downloads in 2011 or the 537 downloads in 2010.

In addition, this year’s challenge had a good balance between student teams and non-student teams (18 of 40).

4.2 Technology

Table 2 summarizes the most commonly used technologies used in VAST Challenge 2012 submissions. This year 30% of the teams used Tableau [2] as part of their submission, which is substantially greater than in previous years. In addition, the D3 [3] and Processing [4] libraries were frequently used by teams who developed custom solutions.

Software Tool	Number of Submissions
<i>Tableau</i>	12
<i>D3</i>	7
<i>MySQL</i>	7
<i>Microsoft Excel</i>	7
<i>Java</i>	5
<i>Processing</i>	5
<i>Postgres</i>	5
<i>SPSS</i>	4
<i>R</i>	3
<i>Other</i>	65

Table 2: Most common technologies used to develop VAST Challenge 2012 submissions

4.3 VAST Papers

All contestants, not just those receiving awards, were welcome to submit a two-page summary paper for the VAST electronic proceedings. As a result 26 of 40 teams who competed this year also submitted a paper.

The following is a list of papers submitted for VAST Challenge 2012, with award titles included if applicable.

Abousalh-Neto, N. A., Kazgan, S., “Big Data Exploration through Visual Analytics.”

Barcelos, Y., Aburjaile, F., Leite, L.R., Oliveira, S.T., “Combining Traditional and High-density Visualizations in a Dashboard to Network Health Monitoring.”

Cao, Y., Moore, R., Mi, P., Endert, A., North, C., Marchany, R., “Dynamic Analysis of Large Datasets with Animated and Correlated Views.” **Challenge 2 Honorable Mention: Good Use of Coordinated Displays.**

Chen, V.Y., Razip, A.M., Ko, S., Qian, C.Z., Ebert, D.S., “SemanticPrism: a Multi-aspect View of Large High-dimensional Data.” **Challenge 1 Award: Outstanding Integrated Analysis and Visualization.**

Choudhury, S., Kodagoda, N., Nguyen, P., Rooney, C., Attfield, S., Xu, K., Zheng, Y., Wong, B.L.W., Chen, R., Mapp, G., Slabbert, L., Aiaash, M., Lasebae, A., “M-Sieve: A visualisation tool for supporting network security analysts.” **Challenge 1 Award: Subject Matter Expert's Award.**

Chung, H., Cho, Y.J., Self, J., North, C., “Pixel-Oriented Treemap for Multiple Displays.”

Dudás, L., Fekete, Zs., Göbölös-Szabó, J., Radnai, A., Salánki, Á., Szabó, A., Szűcs, G., “OWLAP - Using OLAP Approach in Anomaly Detection.” **Challenge 1 Award: Good Support for the Data Preparation, Analysis, and Presentation Process.**

Fischer, F., Fuchs, J., Mansmann, F., and Keim, D.A., “BANKSAFE: A Visual Situational Awareness Tool for Large-Scale Computer Networks.” **Challenge 1 and 2 Award: Outstanding Comprehensive Submission.**

Gibson, H., Vickers, P., “Network Infrastructure Visualisation Using High-Dimensional Node-Attribute Data.”

Harrison, L., Laska, J., Spahn, R., Iannacone, M., Downing, E., Ferragut, E.M., Goodall, J.R., “situ: Situational Understanding and Discovery for Cyber Attacks.”

Hildenbrand, J., Paval, D.I., Thapa, P., Rohrdantz, C., Mansmann, F., Bertini, E., Schreck, T., “VAST 2012 Mini-Challenge 2: Chart- and Matrix-based Approach to Network Operations Forensics.”

Horn, C., Ellsworth, C., “Visual Analytics for Situation Awareness of a Large-Scale Network.” **Challenge 1 Award: Effective Video Presentation.**

Jonker, D., Langevin, S., Schretlen, P., Canfield, C., “Agile Visual Analytics for Banking Cyber ‘Big Data’”

Kachkaev, A., Dillingham, I., Beecham, R., Sarah Goodwin, Ahmed, N., Slingsby, A., “Monitoring the Health of Computer Networks with Visualization.” **Challenge 1 Award: Efficient Use of Visualization.**

Krüger, R., Bosch, H., Koch, S., Müller, C., Reina, G., Thom, D., Ertl, T., “HIVEBEAT - A Highly Interactive Visualization Environment for

Broad-Scale Exploratory Analysis and Tracing.” **Challenge 1 Honorable Mention: Comprehensive Visualization Suite.**

Laberge, L., Kaul, S., Anderson, N., Agnew, C., Goldstein, D., Kolojechick, J., “Enhancing the ‘Think Loop Process’ with Consistent Interactions.” **Challenge 1 Honorable Mention: Good Interaction Techniques.**

Migut, G., van Wees, J., Bakker, D., de Goede, B., Steltenphol, H., Lenferink, N.O., Worring, M., “VAST Challenge 2012: Interactively Finding Anomalies in Geo-temporal Multivariate Data.”

Pabst, R., “BusinessForensics HQ.” **Challenge 1 Award: Good Visual Design.**

Shi, L., Liao, Q., Yang, C., “Investigating Network Traffic through Compressed Graph Visualization.” **Challenge 2 Award: Good Adaptation of Graph Analysis Techniques.**

Shurkhovetskyy, G., Bahey, A., Ghoniem, M., “Visual Analytics for Network Security.”

Stark, R.F., Wollocko, A., Borys, M., Kierstead, M., and Farry, M., “Visualizing Large Scale Patterns and Anomalies in Geospatial Data.” **Challenge 1 Honorable Mention: Good Visual Design.**

Takeda, S., Kobayashi, A., Kobayashi, H., Okubo, S., Misue, K., “Irregular Trend Finder: Visualization Tool for Analyzing Time-series Big Data.”

Williams, F.C.B., Faithful, W.J., Roberts, J.C., “SitaVis - Interactive Situation Awareness Visualization of Large Datasets.” **Challenge 1 Honorable Mention: Good Situation Awareness Snapshot.**

Zhang, T., Liao, Q., Shi, L., “3D Anomaly Bar Visualization for Large-scale Network.”

Zhao, M., Zhong, C., Ciamaichelo, R., Konek, M., Sawant, N., Giacobbe, N.A., “Federating Geovisual Analytic Tools for Cyber Security Analysis.”

Zhao, Y., Zhou, F., Ronghua, S., “NetSecRadar: A Real-time Visualization System for Network Security.” **Challenge 2 Honorable Mention: Interesting Use of Radial Visualization Technique.**

5 PATH FORWARD

This year (2012) has been the “Year of the Contest” where analysis competitions have sprung up around the world, including listings in Challenge.gov, Kaggle, and CrowdAnalytix. The VAST Challenge committee is pleased to have been supporting the visual analytics community with specialized contests to support growth of the science and technology since 2006.

The VAST Challenges are never the same from year to year, but they consistently attract strong interest, both in terms of the number of submissions and the number of dataset downloads. The Challenge is an integral part of the VAST conference, and the committee also takes great pride in receiving entries from all over the world, and then meeting the team members who put so much of themselves into their work at the conference workshop each year.

This year’s workshop is open so that all VisWeek participants can come learn about the Challenge, meet the participating teams, and see the solution software in action. It is hoped that this will generate even more synergy and interest in the Challenge. The committee welcomes comments and ideas from the VAST community to help make the activity increasingly beneficial to all.

ACKNOWLEDGEMENTS

The committee acknowledges the National Security Agency for supporting the VAST Challenge.

The committee also wishes to thank Wendy Cowley, Adam Roberts, Brandon Meador, Cody Tews, John Burnette, and Ian Roberts of Pacific Northwest National Laboratory; Catherine Plaisant at the University of Maryland; and John Fallon and Patrick Stickney of the University of Massachusetts, Lowell.

REFERENCES

- [1] VAST Challenge, <http://www.cs.umd.edu/hcil/vastchallenge/>.
- [2] Tableau Software: <http://www.tableausoftware.com>
- [3] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer, "D3: Data Driven Documents", IEEE Transactions on Visualization & Computer Graphics, IEEE Press, pp. 2301-2309, 2011.
- [4] Casey Reas and Ben Fry, Processing: A Programming Handbook for Visual Designers and Artists, The MIT Press, Cambridge MA, 2007.
- [5] Visual Analytics Benchmark Repository, <http://hcil.cs.umd.edu/localphp/hcil/vast/archive/>.